VOL. 42 1951

4781/58

The Journal of Educational Psychology

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching

EDITORIAL BOARD

STEPHEN M. COREY
Teachers College, Columbia University

JACK W. DUNLAP
10 E. 49th St., New York 17, N. Y.

KARL J. HOLZINGER
University of Chicago

HAROLD E. JONES
University of California

H. H. REMMERS
Purdue University

J. B. STROUD
State University of Iowa

PERCIVAL M. SYMONDS
Teachers College, Columbia University

MILES A. TINKER
University of Minnesota

ALEXANDER G. WESMAN
The Psychological Corporation

PAUL A. WITTY
Northwestern University

H. E. BUCHHOLZ
Managing Editor

Baltimore, Md.
WARWICK and YORK, Inc.
1951

TABLE OF CONTENTS

Academic Achievements of Veterans and Non-Veterans at the City College of New York. LOUIS LAURO AND JAMES D. PERRY 31–42

Analysis of a Procedure Used in the Teaching and Rating of Oral Expression, An. FRANCIS J. DI VESTA AND JAMES H. L. ROACH. 468–478

Awareness of Vocabulary Size: Its Relation to Class Standing and Sex Differences. JANICE LEVIN AND CHALMERS L. STACEY 174–184

Book Reviews. 59–64, 124–128, 190–192, 246–253, 315–320, 372–384, 437–448

Citizenship Attitudes of Graduated Seniors at Purdue University, U. S. College Graduates and High-school Pupils. A. J. DRUCKER AND H. H. REMMERS 231–235

Community Variables as Predictors of Intelligence and Academic Achievement. ROBERT L. THORNDIKE. 321–338

Comparison of Reading Ability and Readability, A. JOHN U. MICHAELIS AND FRED T. TYLER 491–498

Corrective Reading: A Means of Increasing Scholastic Attainment at the College Level. DOROTHY J. MCGINNIS. 166–173

Differential Aptitude Tests as Predictors in Engineering Training, The. RALPH F. BERDIE 114–123

Do Alumni and Students Differ in Their Attitudes toward Instructors? A. J. DRUCKER AND H. H. REMMERS 129–143

Effect of Attention-gaining Devices on Film-mediated Learning, The. D. MORGAN NEU 479–490

Four-phase Learning Project in Teaching Educational Psychology in College, A. D. R. STONE 301–307

Frequent Testing as a Motivating Factor in Large Lecture Classes. MILDRED L. FITCH, A. J. DRUCKER, AND J. A. NORTON, JR.. 1–20

Geo-ethnic Preference Inventory, The: World Cultures and Autistic Thinking. HARRY A. GRACE. . . . 206–214

Hostility, Communication, and International Tensions: III. The Hostility Factors. HARRY A. GRACE AND GLORIA LAUER GRACE. 293–300

How Elementary-school Teachers Meet Selected School Situations. FRANK SLOBETZ. 339–356

Insight into One's Own Values. JULIAN C. STANLEY. . 399–408

Intelligence of College Women as Related to Family Size. AUDREY M. SHUEY AND CARROLL MARTIN HEDRICK. 215–222

Intelligence of Music Students, The. LESTER R. WHEELER AND VIOLA D. WHEELER 223–230

Intrauniversity Variations in Grading Practices. BERNARD M. BASS 366–368

Methodology and the Teaching of Educational Psychology. JOHN E. HORROCKS 277–284

Methods for Indicating the True Size of Pictured Objects. MORTON S. MALTER 409–420

Note on the Equivalence of the Coefficient of Contingency to the Product-moment Coefficient of Correlation, A. C. W. ODELL. 308–310

Notes on a Suggested Index of Item Validity: The U-L Index. A. PEMBERTON JOHNSON 499–504

Notes on the Coefficient of Agreement. J. WILLIAM DUNLAP. 369–371

On Methods of Teaching Elementary Psychology. F. L. MARCUSE. 236–240

One-way Vision and Projection Facilities for Clinic Demonstration and Teaching. HAROLD A. DELP. 311–314

Patterns in the Mass Communications Tastes of the Young Audience. PAUL I. LYNESS 449–467

Personality and Achievement of the Classroom Participant, The. HENRY CLAY SMITH AND DONALD S. DUNBAR. 65–84

Prejudice Verbalization among Children. CONRAD CHYATTE, DOROTHY F. SCHAEFER AND MARTIN SPIAGGIA. 421–431

Problem in Factorial Design, A: The Development of a Picture and Inset Test for Determining Young Children's Responses to Persons of Different Skin Color. CATHERINE LANDRETH AND ELIZABETH FRAZIER PLATT. 385–398

Publications Received. 254–256

Reading Ease, Human Interest Value, and Thematic Content of *St. Nicholas Magazine*, The: A Study of Children's Literature. BARBARA SWARD AND DALE B. HARRIS 153–165

Readability of Children's Textbooks. EDMUND W. J. FAISON 43–51

Relation of Sex, Age and School Achievement to Levels of Aspiration, The. LOWELL M. WALTER AND STANLEY S. MARZOLF 285–292

Relationship Existing between Bilingualism and Racial Attitude, The. GRANVILLE B. JOHNSON, JR.. . . 357–365

Relative Contributions of Vocabulary and an Index of Industriousness for English to Achievement in English. WILLIAM C. KRATHWOHL. 97–104

Relative Effectiveness of Two Types of Responses to Items of a Scale on Attitudes toward Education. CHARLES O. NEIDT AND WILLIAM R. MERRILL . . 432–436

Selection of Students for a Trade and Industrial Education Curriculum. H. S. BELMAN AND R. N. EVANS 52–58

Social Perceptions and Teacher-pupil Relationships. N. L. GAGE AND GEORGE SUCI 144–152

Some Comments Concerning the Rôle of Education in the "Creation of Creation." EDITH A. WEISSKOPF 185–189

Some Sampling Problems in Educational Research. ELI S. MARKS 85–96

Student Project in Educational Psychology, A. GEORGE K. MORLAN 241–245

Studies in Reputation: III. Correlates of Popularity among Elementary-school Children. READ D. TUDDENHAM 257–276

Study of Social Class Motivation, A: Relationships between Anxiety for Education and Certain Socio-economic and Intellectual Variables. A. N. Hieronymus. 193–205

Teen-agers' Attitudes toward Problems of Child Management. H. H. Remmers and A. J. Drucker . 105–113

Validity Study of Personality Questionnaires at the Upper Elementary Grade Level, A. William J. Lodge. 21–30

BOOK REVIEWS

A. Anastasi and J. P. Foley, Jr. *Differential Psychology.* (J. B. Stroud) 60

Bruno Bettelheim. *Love Is not Enough.* (H. Meltzer). . 440

S. H. Britt, Editor. *Selected Readings in Social Psychology.* (J. B. Stroud) 250

Hadley Cantril, Editor, and Mildred Strunk. *Public Opinion, 1935–46.* (A. S. Edwards) 441

Matthew N. Chappell. *In the Name of Common Sense: Worry and Its Control.* (Herbert A. Carroll) . . . 59

Algernon Coleman, Clara B. King, Clare Balluff and Robert H. Fife. *An Analytical Bibliography of Modern Language Teaching.* (Miles A. Tinker) 383

Lee J. Cronbach. *Essentials of Psychological Testing.* (Roger T. Lennon). 124

Loraine A. Dahl. *Public School Audiometry.* (Miles A. Tinker). 64

Allen L. Edwards. *Experimental Design in Psychological Research.* (Leonard S. Kogan) 315

Douglas H. Fryer and Edwin R. Henry, Editors. *Handbook of Applied Psychology.* (N. L. Gage) 372

Regina Mary Goff. *Problems and Emotional Difficulties of Negro Children.* (Philip M. Kitay) 381

J. P. Guilford. *Fundamental Statistics in Psychology and Education.* (J. B. STROUD) 318

August B. Hollingshead. *Elmtown's Youth: The Impact of Social Classes on Adolescents.* (NEAL GROSS). 246

Robert Hoppock. *Group Guidance Principles, Techniques, and Evaluation.* (RALPH F. BERDIE) 126

Mary Alice Mitchell. *The Relationship of Reading to the Social Acceptability of Sixth-grade Children.* (MILES A. TINKER). 252

Norman L. Munn. *Handbook of Psychological Research on the Rat.* (J. B. STROUD). 444

Lawrence I. O'Kelly. *Introduction to Psychopathology.* (H. MELTZER) 318

Gerald R. Pascal and Barbara J. Suttell. *The Bender-Gestalt Test: Qualification and Validity for Adults.* (LEE J. CRONBACH). 438

Jean Piaget and Bärbel Inhelder. *Le Développement des Quantités chez l'Enfant. Conservation et Atomisme.* (WILLIS N. POTTER). 60

Saul Rosenzweig with Kate L. Kogan. *Psychodiagnosis.* (WILLIAM F. HOLMES). 62

Celia B. Stendler. *Children of Brasstown.* (NEAL GROSS). 190

Catherine Stern. *Children Discover Arithmetic: An Introduction to Structural Arithmetic.* (MILES A. TINKER). . . 249

C. P. Stone, Editor. *Annual Review of Psychology.* (J. B. STROUD). 448

Donald R. Taft. *Criminology: A Cultural Interpretation.* (H. MELTZER) 247

University of Chicago, Faculty of the College. *The Idea and Practice of General Education.* (C. M. LOUTTIT). . . . 437

E. G. Williams, Editor. *Trends in Student Personnel Work.* (J. B. STROUD) 191

Werner Wolff. *Diagrams of the Unconscious, Handwriting and Personality in Measurement, Experiment and Analysis.* (H. MELTZER). 445

Arthur Wormhoudt. *The Demon Lover.* (H. MELTZER). . 251

INDEX TO AUTHORS

Bass, Bernard M. 366
Belman, H. S. 52
Berdie, Ralph F. 114
Chyatte, Conrad 421
Delp, Harold A. 311
Di Vesta, Francis J. 468
Drucker, A. J. 1, 105, 129, 231
Dunbar, Donald S. 65
Dunlap, J. William 369
Evans, R. N. 52
Faison, Edmund W. J. 43
Fitch, Mildred L. 1
Gage, N. L. 144
Grace, Gloria Lauer 293
Grace, Harry A. 206, 293
Harris, Dale B. 153
Hedrick, Carroll Martin 215
Hieronymus, A. N. 193
Horrocks, John E. 277
Johnson, A. Pemberton 499
Johnson, Granville B., Jr. 357
Krathwohl, William C. 97
Landreth, Catherine 385
Lauro, Louis 31
Levin, Janice 174
Lodge, William J. 21
Lyness, Paul I. 449
Malter, Morton S. 409
Marcuse, F. L. 236
Marks, Eli S. 85
Marzolf, Stanley S. 285
McGinnis, Dorothy J. 166
Merrill, William R. 432
Michaelis, John U. 491
Morlan, George K. 241
Neidt, Charles O. 432
Neu, D. Morgan 479
Norton, J. A., Jr. 1
Odell, C. W. 308
Perry, James D. 31
Platt, Elizabeth Frazier 385
Remmers, H. H. 105, 129, 231
Roach, James H. L. 468
Schaefer, Dorothy F. 421
Shuey, Audrey M. 215
Slobetz, Frank 339
Smith, Henry Clay 65
Spiaggia, Martin 421
Stacey, Chalmers L. 174
Stanley, Julian C. 399
Stone, D. R. 301
Suci, George 144
Sward, Barbara 153
Thorndike, Robert L. 321
Tuddenham, Read D. 257
Tyler, Fred T. 491
Walter, Lowell M. 285
Weisskopf, Edith A. 185
Wheeler, Lester R. 223
Wheeler, Viola D. 223

478Z58

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 42 January, 1951 Number 1

FREQUENT TESTING AS A MOTIVATING FACTOR IN LARGE LECTURE CLASSES

MILDRED L. FITCH, A. J. DRUCKER, and J. A. NORTON, JR.
Purdue University

Any instructional procedure is useful which stimulates the interest of the student and produces desirable activity leading to educational accomplishment and growth. Although the lecture method, as used in colleges and universities, has not been shown to be an ineffective method of instruction, even with large groups,[13] many instructors, especially in the social sciences, like to supplement lectures with recitation or student-instructor discussions or avoid lectures as much as possible. Large enrollments in basic college courses pretty much preclude discussion or recitation methods whose aim is extensive student participation.

This study deals with the investigation of two interrelated instructional devices employed to improve student achievement in a lecture course in Government, where classes number from one hundred to two hundred. To the regular three one-hour lecture sessions are added several hour-long discussion sessions, attendance at any of which is optional on the part of the student. The other device is a program of weekly quizzes over course material, where results of such testing are not used in determining the course grade. Frequent measurement is expected to result in steadier application of the individual to the task at hand, this instructional function of measurement being best served when divorced from the regular process of achievement evaluation.

There is some objective evidence that frequent testing in the classroom results in better achievement.[4,12] Knowledge of results, a probable outcome of frequent testing, has also been shown to aid learning.[1,8]

EXPERIMENTAL PROCEDURE

The current study was undertaken to explore further the effects of frequent testing upon motivation of college students to achieve and the students' outside endeavor with respect to a particular course content. The achievement of two large classes of students in Government 10, a required course in the School of Science at Purdue University, were used to determine the benefits to learning gained when (a) short quizzes for the purpose of guiding the student's own achievement were given in one class over the weekly assignments, and (b) voluntary discussion groups were provided for both classes each week in addition to the three regular class periods. The control group was designated as the class of ninety-seven students to whom only the regular monthly quiz was given, and the experimental group consisted of the other class of one hundred ninety-eight students to whom the weekly quiz was administered in addition to the regular monthly quiz. Both classes met at 9 o'clock on alternate days three times a week. The senior author was the lecturer for the course. Assignment of individuals to these classes was not believed to be related to factors of sex, age, ability or curriculum. No knowledge of the study was given to the students.

The lecture method was used over identical material for both classes and the same reading assignments were made. The last half hour of the third class meeting each week was open for discussion or questions from the floor, primarily over the reading assignments. This was followed by a ten-minute quiz in the experimental section only, and this quiz was confined to the reading assignments in the texts and not to the supplementary material covered in the lectures. Thus both classes had identical lectures and reading assignments and opportunity to raise questions for discussion in one regular class period.

Four one-hour tests were administered to both groups during the semester over the work covered each month. A fifth test, which reviewed topics covered during the semester, was also given at the end of the course. The first, fourth and fifth tests were objective—true-false, multiple choice, matching or short answer—and the second and third were essay-type tests.

The three objective tests were identical for both classes. The two essay tests were composed of 'thought' questions given to

478 Z 30

students of both classes before the day of testing. Every effort was made to make the essay tests comparable. All tests were administered to the experimental section before the control section; hence it is believed that advance knowledge of the objective test questions obtained before the test would benefit only the control section.

Six voluntary discussion sections, set up at different hours during the middle of the week at times which would best accommodate the students, were handled by two graduate students.* These sections were arranged as an outlet for discussion for students who felt they would learn more if they had the opportunity for more personal contact with an instructor than large lecture sections could provide. It was evident, however, that attendance at the discussion sections was increased when topics for monthly test review could be discussed, although this was not the original intention. A poll of the sections showed that the use of the discussion groups on a voluntary basis was favored by a substantial majority of both groups. Weekly quizzes were also favored by both groups, the control people having become familiar with weekly quizzes through their use in Government 9 the previous semester. As a matter of fact, when the weekly quizzes were not resumed at the start of the semester for the control section, some complaints were registered by students who had favored that technique in Government 9 as a device to encourage them to read their weekly assignments.

ANALYSIS AND RESULTS†

The Variables.—The criterion variable for evaluating the experimental results is based on the grades assigned in the five one-hour tests given in both experimental and control sections, and graded jointly for both sections. The grades assigned on each test were based on the Purdue University number grading system, with values one through six, but greater differentiation was available through the use of plusses and minusses appended to the number grade. These grades were assigned on each test on the basis of the frequency distribution of scores on that test, and were determined by a standard grading curve giving the approximate proportion of students who should receive each grade. Thus the grade distribution on each test represents not only a standardized, but also, approximately, a normalized variable, since the standard grading curve was based (roughly) on the normal curve. The criterion variable Y was based on the sum of the grades on these five tests. A sum of approximately normal variables, Y itself was found to be approximately normally distributed.

---

* Mrs. Ingebord Vance and Mrs. Leone Katz also assisted in the compilation of the data and preliminary statistical work.

† By Mr. Norton.

Because of the conditions of scheduling, registration, and other administrative processes which make it difficult to carry on a controlled educational experiment in a large university situation, the experimental and control sections cannot be represented as random samples from any definable common population. For these reasons it was desirable to exercise statistical control over the nature of the two groups to be used for evaluating the experimental variable. The most easily available, as well as the most relevant variable for this purpose was the grade received in the preceding semester of the same course, Government 9. This grade is designated as X, the independent or control variable. It was available only in the gross number grades of the Purdue University grading system, with six possible values, and only five occurring in the sample used for analysis. However, this variable is believed to be entirely satisfactory for the purposes of statistical control for which it is used here. An additional valuable property of the criterion variable Y is that the regression of Y on X was almost perfectly linear.

Factors for the Analysis.—Table 1 gives in percentage form the distribution of frequency of voluntary attendance at the discussion groups for all members of both sections whose records were sufficiently complete to make them eligible for inclusion in the analysis. Table 1 also gives the break-down into four categories which was subsequently used for the discussion attendance variable, and the percentage distribution for this break-down. The upper two boundary lines for this break-down were chosen to conform, approximately, with natural valleys in the two distributions.

The distributions in Table 1 invite interesting speculative hypotheses. In the first place, an almost identical percentage of cases in each section attended no discussion groups. One is tempted to speculate that these represent the students who are

TABLE 1.—ATTENDANCE AT DISCUSSION GROUPS: ALL ELIGIBLE CASES IN BOTH SECTIONS

Discussion Groups Attended	Control Section			Experimental Section		
	N	Per cent	Per cent for grouping chosen	*N*	Per cent	Per cent for grouping chosen
14	0	0	16 = 17.6%	11	5.9	76 = 40.9%
13	1	1.1		14	7.5	
12	1	1.1		8	4.3	
11	1	1.1		10	5.4	
10	5	5.5		12	6.5	
9	3	3.3		9	4.8	
8	4	4.4		8	4.3	
7	1	1.1		4	2.2	
6	5	5.5	19 = 20.9%	3	1.6	21 = 11.3%
5	5	5.5		2	1.1	
4	5	5.5		9	4.8	
3	4	4.4		7	3.8	
2	10	11.0	23 = 25.3%	4	2.2	21 = 11.3%
1	13	14.3		17	9.1	
0	33	36.3	36.3%	68	36.6	36.6%
Total	91	100	100%	186	100	100%

simply not interested in any activity in connection with the course which is not required of them, and the equal proportions suggest that this attitude is in no way affected by a motivational device of the type experimented with here (frequent quizzing). Of the sixty-four per cent of the students who would and did attend one or more voluntary discussion groups, however, there was a marked difference in behavior between the two sections. Of the students in the control section who would attend any discussion groups at all, only about one-quarter attended more than six. In all probability this is directly related to the fact that these students were given only the five one-hour tests during the course of the semester, as it has been noted above that the size of the attendance at the discussion groups seemed to be related to the dates of these tests. In contrast, of the students in the experimental, frequently-quizzed section who would attend any discussion groups at all, nearly two-thirds attended seven or more. It seems clear that for those students who will participate at all in a voluntary activity of this sort, the motivational device experimented with here bears a significant and important relationship to frequency of attendance at the discussion groups.

Several questions now arise. Is attendance at discussion groups related to achievement in the course, as measured by Y? If it is, and if the experimental section shows superior achievement over the control section, to what extent can this superiority be explained by the more frequent attendance at discussion groups motivated in the experimental section? Even more important, what are the answers to these same questions if the effects of the variable X (Government 9 grades) which, among other things, presumably measures initial interest and ability in this course, are partialled out? The analysis was designed to answer these questions.

Design of the Analysis.—Since it was desired to make the comparison between the achievement of the experimental section and that of the control section taking into account the factor of discussion attendance, it was necessary to classify the data according to a double-classification scheme. Table 2 shows this scheme as a 4 × 2 table of eight cells: four classes on the vertical axis representing the four groupings of discussion attendance frequency given in Table 1 and two classes on the horizontal axis representing membership in the control or in the experimental section.

The methods employed were those of analysis of covariance. Our application of these methods differs, however, from those more commonly occurring. We have pointed out, in the discussion of Table 1, the marked dissimilarity in the distributions into discussion attendance classifications between the control and experimental sections. This is a situation that is frequently called the case of disproportionate frequencies (the 'non-orthogonal case'). The disproportionate frequencies situation requires special procedures which we shall not go into in detail. (See Snedecor[9] or Tsao.[11]) We shall, however, discuss some of the basic considerations involved.*

In the case of disproportionate frequencies one must choose between two basic assumptions on which to proceed:[11]

Assumption A: We are interested in making inferences for a population which has equal frequencies in the cells formed by our classificatory factors; the unequal and disproportionate cell frequencies in our sample are merely a result of chance sampling effects, and are not characteristic of the population in which we are interested.

Assumption B: The disproportionate cell frequencies in our sample are the result of characteristics inherent in the classificatory factors to be studied, and we are interested in inferring to a population whose cell frequencies are stratified approximately as are those in our sample.

From our discussion of the dissimilar distributions in Table 1 it is clear that Assumption A was not the one for us. We chose to assume instead that the dissimilarity in the two distributions in Table 1 was a direct result of the motivational device which constitutes the experimental teaching method, and this is certainly just as true for the population to which we are interested in inferring as it is for the sample at hand. We could not, however, make Assumption B and proceed on the basis of the distributions as given in Table 1, either. To do so would have meant that we were assuming that it is always characteristic of the experimental, frequently-quizzed groups to run, on the average, approximately twice as large as less frequently-quizzed groups like the control group! Clearly we did not want to include scheduling irregularities as part of our experimental hypothesis.

---

* No attempt will be made here to go into detail on the theory and procedures of these methods. A word is in order, however, on the intent and function of the analysis of covariance in an investigation of this sort. We have mentioned that we desired to exercise statistical control over the variable X—final grade in Government 9. In most research work in education and other social sciences, such control is achieved by matching cases in the two groups to be compared on the variable to be controlled. If this is done *ex post facto*, as it frequently must be,[2] there is usually a large resulting loss in cases which must be discarded because no match can be found for them in the other group or groups. The very great advantage of the methods of analysis of covariance in an investigation of this sort is that it achieves the effect of such individual matching on the variable to be controlled, without the necessity of carrying out the actual matching procedures, and therefore without the loss of cases that usually attends such procedures. Furthermore, the application of these methods is readily extended to more highly complex designs, involving multiple-classification on several factors to be studied simultaneously (as in our case, where we have a design involving double classification on two factors to be evaluated simultaneously), whereas individual matching procedures become prohibitively unwieldy on such complex designs. The methods also extend to the statistical control of more than one variable, where again individual matching procedures become difficult, if not impossible. For additional information about covariance see references 6, 5, 3, and 9.

This problem was resolved as follows: using a table of random sampling numbers (5) or (9), a random sample of ninety-one cases was selected from the one hundred eighty-six eligible cases in the experimental section. Using these ninety-one cases, and the ninety-one eligible cases from the control section, we could then set up our table of double classification for analysis. This is given in Table 2, which shows equal total membership in the control and the experimental sections, thus obviating the difficulty referred to in the last paragraph, but nevertheless retains, within random sampling fluctuations, the dissimilar distributions of discussion attendance in the two sections. On such a table we were able to proceed under Assumption B.

A study of Table 2 will be helpful in understanding some of the results of the later analysis. We observe, first, that the experimental section with a mean of 62.7 does seem to exceed substantially the control section with a mean of 55.5 on the criterion variable Y. However, note that the experimental section also has a greater mean on the initial variable X than does the control section. This latter difference might substantially affect the comparison between the two groups on Y. Next note the pattern of the marginal X-means for the discussion attendance groupings. For those students who attended any discussion groups at all in this course, the average final grade in the preceding semester of the same course decreases regularly with decreasing attendance at such discussion groups. This would seem to lend support to the hypothesis that the X variable serves as a measure of interest in and/or conscientious endeavor with respect to this course, which, for these students, led them to attend one or more of the discussion groups. The students who attended no discussion groups, however, reverse this trend, and have an X-mean not far from the grand X-mean for all cases. This seems to support the speculation, made above, that these students differed in some characteristic qualitative manner from the rest, whose observed effect is that they simply did not (perhaps could not) attend voluntary discussion groups, but that this does not necessarily imply lower interest and/or ability in this course than that displayed by students who did attend discussion groups. Because the two sections were not comparable on the X variable, and because the X variable seems to be related to discussion attendance for all those above the 'None' group, it is of particular

TABLE 2.*—DOUBLE CLASSIFICATION TABLE: ALL ELIGIBLE CASES FROM CONTROL SECTION; A RANDOM SAMPLE OF EQUAL N FROM EXPERIMENTAL SECTION

			Control Section Infrequent Quizzing		Experimental Section Frequent Quizzing		Total
Discussion Attendance	7–14	N	16	(17.6%)	34	(37.4%)	50
		M_x	4.31		4.21		4.24
		M_y	63.6		65.6		65.0
	3–6	N	19	(20.9%)	14	(15.4%)	33
		M_x	4.05		4.36		4.18
		M_y	59.1		62.6		60.6
	1–2	N	23	(25.3%)	10	(11.0%)	33
		M_x	3.70		3.9		3.76
		M_y	50.6		60.5		53.6
	None	N	33	(36.3%)	33	(36.3%)	66
		M_x	3.88		4.15		4.02
		M_y	53.0		60.4		56.7
Total		N	91	(100%)	91	(100%)	182
		M_x	3.95		4.18		4.06
		M_y	55.5		62.7		59.1

* In the case of this and all subsequent tables, calculations were carried to several more significant digits than those given.

importance that we propose to 'partial out' by statistical control the 'predetermination' effects of X in making our comparisons. Note, finally, that the percentage distribution into discussion attendance groupings in the experimental section in Table 2 is similar to, but not exactly the same as that in Table 1. This, of course, is due to the sampling process in which we have engaged in order to obtain the ninety-one cases which represent the experimental method in Table 2. The largest discrepancy between the percentages in Table 1 and those in Table 2—that in the 3–6 discussion attendance group—is, however, no larger than might be expected by chance, with a t-value of 1.7 for the comparison between the proportion for the ninety-one cases in Table 2 and that for the complementary omitted group of ninety-five. This, of course, is to be expected, since the selection process was a random one.

We have been discussing at some length considerations involved in the double classification design used here because this was necessary to set the stage for the choice of the sample actually used in the analysis, and also helps to understand what the design is intended to accomplish. The first steps in the analysis, however, did not involve the double classification scheme, but dealt with the two factors separately, as single classifications. That is, we compared the achievement of the experimental with that of the control section, ignoring the distribution into discussion attendance groupings. Likewise, we studied the variation in achievement of the four discussion attendance groupings, ignoring membership in the experimental or control sections. Then, as the final step in the analysis, we dealt with the table of double classification, comparing the achievement of the experimental with that of the control section, taking into account the distribution into discussion attendance groupings. And *vice versa*. Comparison of the results in the single and double classification analyses then helped answer some of the questions we posed at the end of the previous section.

Assumptions Underlying the Analysis.—Before passing to the discussion of the analysis itself, it would be well to discuss briefly the principal assumptions inherent in the application of the methods of analysis of covariance:

1) That the regression of the criterion variable on the variable to be controlled (Y on X in our case) is fundamentally the same in each of the several cells.

2) That this regression is linear.

3) That the residuals after partialling out the effects of X are fundamentally normally distributed in each of the several cells.

4) That these several normal distributions of residuals all have the same variance (assumption of homogeneity of residual variance).

5) That the cases in our sample are random samples from each of these populations.

Not all of these assumptions are equally crucial. There is empirical evidence, for instance, that 'moderate' departure from the normality assumption does not seriously affect the validity of the methods. The most crucial assumption is considered to be that involving homogeneity of residual variance. We have already mentioned that, for the total sample used for analysis, the variable Y was approximately normally distributed, and the regression of Y on X quite satisfactorily linear. This, of course, tells us nothing about the situation in each of the several cells of our double classification or about the residuals, but it lends plausibility to the assumption that the same conditions hold in each of the cells without resorting to an explicit test of these assumptions.

Because the assumption of homogeneity of residual variance is considered to be most crucial, we performed an explicit test of this assumption based on the double classification design, since our most important conclusions were to be based on the results of the analysis of covariance in the double classification design. Using Hartley's M-test, with the tables by Thompson and Merrington,[10] the obtained M was found to be well within the five per cent level of significance for M, and the homogeneity assumption therefore tenable. Furthermore, homogeneity of residual variance in the eight cells of the table of double classification implies homogeneity of residual variance in the groups of each of the single-classification analyses performed first, since each of these groups is but a combination of some of the eight cells.

The Single Classification Analyses.—Table 3 presents the two single-classification analyses—that on method, ignoring discussion attendance, and that on discussion attendance, ignoring method. We have seen that the experimental group had a Y-mean appreciably higher than that for the control section, but we have also seen that there was a difference between the two groups in the same direction on whatever student characteristics are measured by the variable X, which presumably include interest and/or ability in this Government course. To what extent was the difference between the methods groups on the criterion variable Y due to the predetermination effect of this non-comparability in X, or rather, to what extent was it not due to these predetermination effects, and therefore, by implication, probably was due to the difference between the instructional methods used in the two sections? The answer is contained in Columns 11 and 12 of Table 3. There we see that even with the effects of the X-variable partialled out, the F-ratio is highly significant, with a probability of arising purely by chance alone only somewhere between ½ and ⅒ of one per cent. We may conclude that, barring systematic effects that could bring about this result from variables not controlled in this study, the experimental, frequent-quizzing method did bring about significantly greater achievement on the part of those students to whom it was applied. Note that nothing is said in this conclusion about how much, if any, of this superiority is due to the greater attendance at discussion groups which, we have noted, was apparently motivated by the experimental, frequent-quizzing method. This question could not be answered until we did the double-classification analysis.

Column 13 gives correlation coefficients which measure the degree of the relationship between the X and Y variables. In the row for 'Total' appears the coefficient which is the one we usually have in mind when we speak of the correlation between X and Y over this sample of one hundred eighty-two cases. However, we have noticed that the experimental section had both a higher X-mean and a higher Y-mean than did the control section—that is, some of this 'total' correlation is associated with group membership. The coefficient in the row for 'Within Methods Groups' is a correlation coefficient which measures the extent of the relationship between Y and X that is not associated with group membership. It is called, appropriately, the 'within-groups correlation,' as opposed to that portion of the total relationship which is associated with 'between-groups' differences. That it is slightly smaller than the total correlation coefficient indicates, as we have already noted, that some small portion

TABLE 3.—ANALYSES FOR SINGLE CLASSIFICATIONS: METHODS AND DISCUSSION ATTENDANCE

(Col.) (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Source	df before adj.	Σx^2	Σxy	Σy^2	Adj. Σy^2	df for Adj. Σy^2 (also for r)	Reduced Σy^2	df for Reduced Σy^2	$MS_{y \cdot x}$	$F_{y \cdot x}$	Probability	r_{xy}
Method (Quiz frequency)	1	2.42	75.1	2329			976	1	976	8.94	$.005 > P > .001$	
Within Methods Groups	180	151.91	1635.9	37167	19551	179			109			.688
Total	181	154.34	1711.0	39495	20527	180						.693
Discussion Attendance	3	5.26	121.0	3187			1177	3	392	3.59	$.025 > P > .01$	
Within Discussion Attendance Groups	178	149.07	1590.0	36309	19351	177			109			.683
Total	181	154.34	1711.0	39495	20527	180						.693

of the total relationship is associated with between-groups differences.

The second portion of Table 3 gives the analysis for discussion attendance groupings as a single classification. We have seen in Table 2 that there is considerable variation in X-means for the discussion attendance groupings. The question arises again: To what extent is the variation in achievement among the several discussion attendance groupings, as measured by the criterion variable Y, not due to any predetermination effects of the student characteristics measured by X? The answer is contained in columns 11 and 12 where we see that, even after partialling out the effects of X, the variation in achievement among the discussion attendance groupings is moderately large, though not quite significant at the one per cent level, the probability for the obtained F having arisen by chance alone being somewhere between two and one-half and one per cent. We may conclude as follows: there was a tendency for students who differed on characteristics measured by X, which, as we have said, presumably include interest and/or ability in this course, to differ in their frequency of attendance at the voluntary discussion groups; when we have removed the effects of this tendency, it appears that frequency of attendance at discussion groups is still related to achievement in the course, though the relationship is only moderate, failing slightly to meet the one per cent level of significance.

Column 13 is interpreted as indicated above. Note that the within-groups correlation falls farther below the total correlation than was the case in the methods analysis. This implies that more of the total relationship is associated with variation among discussion attendance groupings, than was the case for method-group differences.

Let us now summarize the conclusions we have been able to draw so far from our experimental data. Even when the predetermination effects of the student characteristics measured by X are partialled out, we have seen that there is still a highly significant difference in achievement in the course between the experimental and the control sections, and we have also seen that attendance at the discussion groups bears a relationship to achievement in the course. Now, however, we have observed in Tables 1 and 2 that discussion attendance behavior differed markedly between the two sections—the 'attenders' in the experimental, frequently-quizzed section were motivated to attend many more discussion groups, on the average, than were the attenders in the control section. The following question now arises: how much of the significant superiority in achievement of the experimental section over the control section can be explained by the greater attendance at discussion groups which was motivated in the experimental section? Can all of it be so explained? If not, how large (how significant) is that part which cannot be so explained? The double classification analysis to follow undertook to answer these questions.

*The Double Classification Analysis.**—The double classification analysis is presented in full in Table 4. Here the tests of the difference between the two methods groups are tests taking into account the variation in discussion attendance. Similarly, the tests of variation among discussion attendance groups are tests taking into account method-group membership.

Columns 11 and 12 give the significance tests on variation in achievement after partialling out X. We see that the F-ratio for method has been drastically reduced from that in the single classification. This reduction is due to the fact that attendance at discussion groups is now being taken into account in the test of differences between the methods groups; the methods groups differed in their discussion attendance behavior, and discussion attendance was related to achievement in the course (Table 3). Similarly we see that the F-ratio for variation among discussion attendance groups is reduced over that in Table 3.

The result of greatest interest to us, however, is that, even when we take differing discussion group attendance into account, and also partial out X, there still remains a superiority of the experimental group over the control group of considerable mag-

---

* We shall not go into a detailed explanation of a new factor introduced in the double classification design—the so-called 'Interaction' effect between method and discussion attendance, because it plays no important rôle in our results. It must be included, however, in order to make the analysis complete. In fact, in the analysis of Table 4 the Interaction effect was clearly insignificant, and was combined with 'within-cells' to yield a combined 'residual' error estimate to be used in testing the main effects of method and attendance at discussions. We are therefore dealing only with the two main effects in Table 4.

TABLE 4.—COMPLETE ANALYSIS: DOUBLE CLASSIFICATION, METHOD x DISCUSSION ATTENDANCE; NON-ORTHOGONAL CASE, ASSUMPTION B

(Col.) (1) Source	(2) df before Adj.	(3) Σx^2	(4) Σxy	(5) Σy^2	(6) Adj. Σy^2	(7) df for Adj. Σy^2 (also for r)	(8) Reduced Σy^2	(9) df for Reduced Σy^2	(10) $MS_{y \cdot x}$	(11) $F_{y \cdot x}$	(12) Probability	(13) r_{xy}
Method (Quiz freq.)	1	1.23	40.9	1364			636	1	636	5.98	$.02 > P > .01$	
Discussion Attendance	3	4.07	86.8	2222			836	3	279	2.62	$.10 > P > .05$	
Interaction	3	1.16	12.7	364			226	3	75	0.70	$.75 > P > .50$	
Within Cells	174	146.69	1536.3	34580	18489	173			107			.682
Total	181	154.34	1711.0	39495								.693
Interaction + Within Cells = Residual	177	147.85	1549.0	34944	18715	176			106			.681

nitude, the obtained F-ratio having a probability of arising by chance only somewhere between two and one per cent. That is to say, when we remove the predetermination effects measured by X, and when we remove the effect of disproportionate distribution of discussion attendance, there still remains a superiority in achievement of the experimental section over the control section that barely falls short of being significant at the one per cent level.

We are now in a position to answer the question which we posed ourselves at the end of the preceding section. It is clear, first, that frequent quizzing did bring about greater achievement in this course. But we can now say, further, that while some of this greater achievement can be explained by the greater attendance at discussion groups (on the part of attenders) which the frequent quizzing apparently motivated, by no means all of it can be so explained; there remains a superiority which cannot be so explained that falls between the two and the one per cent levels of significance.

Indeed, we can do more; we can trace the principal source of this remainder. In Table 5, in the column headed $W_rD_{r_y}$ we see that by far the largest contribution to $\Sigma W_rD_{r_y}$ comes from the non-attenders, and the next largest contribution comes from the low attenders in the 1–2 group. This means that the major portion of the methods difference which remains significant in the double classification of Table 4 is due to these two groups. This may be interpreted as follows: the high attenders in the experimental section had high achievement, but they were expected

TABLE 5.—DOUBLE CLASSIFICATION, METHOD X DISCUSSION ATTENDANCE: EXTRACTED PORTION FROM TABLE FOR ESTIMATION OF SUMS OF SQUARES AND PRODUCTS

Discussion Attendance	$W_r = \dfrac{N_{r1}N_{r2}}{N_{r\cdot}}$	$D_{r_y} = \bar{Y}_{r2} - \bar{Y}_{r1}$	$W_rD_{r_y}$
7–14	10.88	+1.96	+21.4
3– 6	8.06	+3.47	+27.9
1– 2	6.97	+9.93	+69.2
None	16.5	+7.39	+122.0
Total	42.41		+240.5

to have high achievement; their counterparts in the control section also had high achievement, so that the high attenders contribute little to the methods difference in the double classification of Table 4 (although they contributed a great deal to the methods difference in the single classification of Table 3, where discussion attendance was not taken into account, because there were a great many more of them in the experimental section than there were in the control section); on the other hand, low- and non-attenders in the experimental section apparently achieved a great deal more than their counterparts in the control section, and it is these groups that account for the significant methods difference that remains in Table 4. We are thus led to speculate that the frequent quizzing method, besides motivating greater attendance at discussion groups on the part of attenders, also motivated greater *outside* endeavor with respect to the course, which we have no measure of in this study; moreover this greater outside endeavor shows up dramatically in the case of the low- and non-discussion group attenders, which accounts for the significant methods difference still remaining when discussion attendance is controlled in Table 4.

It is of considerable interest, from the point of view of statistical methodology, to note that the question which we posed ourselves at the end of the last section, and have now answered, could not have been answered without resorting to the non-orthogonal case, Assumption B. Even those research workers who occasionally use the non-orthogonal case tend to look upon it as an inconvenient nuisance, made necessary by bad luck, or bad planning, or both, that would have been better avoided. This may frequently be true, especially when Assumption A is the appropriate one (that unequal cell frequencies are merely chance sampling effects). But here we had a situation where the disproportionate frequencies are clearly characteristic of the classificatory factors under study, and proceeding under the non-orthogonal case, Assumption B, we have been able to answer a question about our experimental results which could not otherwise have been answered.

Since the test of discussion attendance variation in Table 4 is not of major interest in drawing conclusions about our experiment, we shall not discuss it in detail, but shall only note that, when group membership is controlled, as well as X partialled out, we see in Column 12 that the variation among discussion attendance groups drops below the five per cent level of significance.

Column 13 may be interpreted as previously indicated.

SUMMARY AND CONCLUSIONS

A non-orthogonal case (disproportionate frequencies) of analysis of covariance is demonstrated in the investigation of an independent educational variable in relation to another variable which is differentially distributed in the experimental and control groups.

Frequently-quizzed students in Government have significantly higher achievement than students receiving only monthly quizzes, even after a predetermining index of ability in Government has been partialled out.

This superiority is accounted for only in part by voluntary attendance at discussion groups in Government conducted in addition to class lectures. Even when discussion attendance as well as predeterminers of ability are taken into account, the superiority of the frequently-quizzed students is a significant one. Much of this superiority is attributed to the greater achievement of frequently-quizzed students who attended no or few discussion groups over the achievement of non-frequently-quizzed students who attended no or few discussion groups.

Additional evidence suggests that frequent quizzing motivates some students to attend more discussion groups.

On the basis of these findings we may conclude that frequent testing of achievement in the college lecture classroom may motivate such outside endeavor as will result in superior achievement. The instructor who wishes to use such motivational learning devices to accompany lectures would do well to make available instructional supplements such as the extra discussion groups of this study or other instructional materials and experiences closely correlated with course content.

BIBLIOGRAPHY

1) W. F. Book and L. Norvell. "The will to learn," *Pedagogical Seminary*, 1922, 29, 305–362.

2) E. Greenwood. *Experimental Sociology*, New York: Kings Crown Press, 1945.

3) R. W. B. Jackson. *Application of the Analysis of Variance and Covariance Method to Educational Problems*, Bulletin 11, Dept. of Educ. Research, University of Toronto, 1940.

4) N. Keys. "The influence on learning and retention of weekly as opposed to monthly tests." *J. Educ. Psychol.*, 1934, 25, 511–20.

5) E. F. Lindquist. *Statistical Analysis in Educational Research*, Boston: Houghton-Mifflin, 1940.

6) Q. McNemar. *Psychological Statistics*, New York: John Wiley & Sons, 1949.

7) Maxine Merrington and Catherine M. Thompson. "Tables of percentage points of the inverted Beta (F) distribution," *Biometrika*, 1943, 33, 75–88.

8) L. Panlasigui and F. B. Knight. "The effect of awareness of success or failure," *Twenty-Ninth Yearbook of the National Society for the Study of Education*, Part II, 1930, 611–619.

9) G. W. Snedecor. *Statistical Methods.* 4th Ed. Ames, Iowa: Iowa State College Press, 1946.

10) Catherine M. Thompson and Maxine Merrington. "Tables for testing the homogeneity of a set of estimated variances," *Biometrika*, 1946, 33, 296–304.

11) F. Tsao. "General solution of the analysis of variance and covariance in the case of unequal or disproportionate numbers of observations in the subclasses," *Psychometrika*, 1946, 11, 107–128.

12) A. H. Turney. "Effect of frequent short objective tests upon the achievement of college students in Educational Psychology," *Sch. & Soc.*, 1931, 33, 760–762.

13) *Encyclopedia of Educational Research. The American Educational Research Association.* New York: The Macmillan Co., 1941, p243.

A VALIDITY STUDY OF PERSONALITY QUESTIONNAIRES AT THE UPPER ELEMENTARY GRADE LEVEL*

WILLIAM J. LODGE
Professor of Education Chico State College

THE PROBLEM

This study appraises the validity of a direct and an indirect form of a personality and attitude questionnaire at the fifth-, sixth-, and seventh-grade levels, as measured by correlations with external criteria and by an analysis of responses to the individual items. The study also offers evidence bearing on the effects on questionnaire responses at these grade levels of the requirement or non-requirement of signatures.

The experiment was designed to evaluate and extend a study by Ellis[1] in which two forms of direct and two forms of 'equivalent' indirect questions were combined into one measuring instrument in an effort to determine which form could best distinguish between groups of 'normal' and 'problem' seventh- and eighth-grade boys. None of the four forms did so distinguish, but the indirect type used in the present investigation was regarded by Ellis as the most promising. Examples of the two types employed in the present study are as follows:

Direct: I cry
Very often Pretty often Seldom Never
Indirect: Children who often cry are
Very queer Pretty queer A little queer Not at all queer

Ellis'[1] theory, (4) which is not confirmed by the present experiment, is that children who themselves possess the characteristic of frequent crying will classify other children who often cry as 'Not at all queer,' and that, in effect, the child, while ostensibly rating other boys and girls, will actually be revealing his own characteristics.

EXPERIMENTAL PROCEDURE

The questionnaire employed by the writer consists of forty-three items, thirty-one of which were included in the total of

---

* Summary of a Ph.D thesis on file in the University of California Library, Berkeley, entitled *A Study of Some Factors Affecting Responses on Personality Questionnaires.*

11.8.70
999

thirty-six used by Ellis.[1]* They were obtained by him from a frequency count of the items employed in eight popular personality inventories. Five Hartshorne and May[3] 'ringer' items† and seven cheating attitude questions based on a review of the literature were added by the writer. The foregoing forty-three items were drawn in random order and were then written in direct and indirect form respectively, in the order drawn by chance. Split-half Pearson r's based on one hundred and one cases for the direct form and one hundred and five for the indirect were .83, P.E. = .02 and .86, P.E. = .02, respectively.

The comparison between the two forms of the questionnaire is based on the administration of the instrument by the writer to two groups of two hundred forty-four children each, matched for grade, sex, IQ, reading comprehension and socio-economic status. The subjects were drawn from five California public schools. One-half of the children in each group were tested with signatures required and the other half anonymously, with identification subsequently accomplished by the experimenter by means of a pinprick code system. The questionnaires were administered a second time one week later by the classroom teachers.

The 'personality' items on the questionnaire were scored in accordance with the system developed empirically by Ellis[1]; i.e., one point was scored on the direct form for each 'Very often' or 'Pretty often' answer and no points for each 'Seldom' or 'Never' response. The indirect form was scored one point for each 'Not at all queer' reply and no points for the 'Very queer,' 'Pretty queer,' and 'A little queer' answers. The cheating attitude questions were scored in the same way except that 'Seldom' answers on the direct form were also scored one point, since they obviously constitute a cheating admission.

A social status score for each child was determined by the administration on the fourth day of a sociometric instrument by the classroom teachers. This instrument employed the three criteria used by Flotow,[2] involving up to three choices of children with whom the child would like to work and play and next to whom he would like to sit. Flotow reports, in non-quantitative terms, a high degree of validity and reliability. The directions to the children followed closely the pattern recommended by Moreno[4] to obtain maximum validity.

---

* By permission of the American Psychological Association, Inc., and Psychological Monographs.

† By permission of the Macmillan Company.

Each child's social status score as employed in this study was determined by calculating the total number of times he was chosen by others, with specific criterion and one-two-three order of choice disregarded.

The *Van Wagenen Unit Scales of Achievement: Reading Comprehension Tests*[6] were administered on the twenty-first day for the dual purpose of obtaining standardized reading comprehension scores for matching and correlational purposes and to control a classroom cheating experimental situation. The papers were scored by the experimenter with no marks recorded on them and then returned promptly to the children, ostensibly for self-scoring. A cheating behavior rating was subsequently determined for each child by dividing the difference between his true score and a perfect score into the number of points he raised his score, if any.

Finally, socio-economic ratings were determined by converting parent's occupation in accordance with the *Barr Scale Ratings of Occupational Status*[5] (66–72).

STATISTICAL FINDINGS

Table I shows that the direct and the indirect forms of the questionnaire are significantly different according to a rigid chi-square test which was used because of the skewed distribution of the data. Correlations of various sections of the questionnaire, however, with external criteria (Table III) suggest that neither form possesses much validity as measured. The r's in Table III indicate some superiority for the indirect form, but positive r's between personality problems as scored by the indirect form and intelligence, reading comprehension, and socio-economic status (Table IV) are interpreted as meaning that, if the theory held by Ellis[1] and by others is tenable, superior children tend to have a greater number of personality problems than do those below the mean. It might be maintained that the more intelligent children are more self-critical, which would be reflected in higher 'neurotic' scores, but neither the literature generally nor the item analysis (Tables V, VI) particularly supports such an hypothesis.

The data suggest that many children do not, in effect, rate themselves when they respond to most of the indirect items but do

TABLE I.—COMPARISON OF THE DIRECT FORM OF PERSONALITY QUESTIONNAIRE WITH THE INDIRECT FORM: UNFAVORABLE RESPONSES

Category	Direct		Indirect		Chi Square of Difference	P	Significance Level
	N	Mean	N	Mean			
Personality Score, 1st Testing*	244	6.20	244	11.93	64.24	.01	1 per cent
Personality Score, 2nd Testing*	244	5.29	244	12.64	65.01	.01	1 per cent
Cheating Attitudes, 1st Testing	244	1.79	244	.95	6.74	.05	5 per cent
Cheating Attitudes, 2nd Testing	244	1.90	244	1.05	8.00	.02	5 per cent

* Includes the following sub-categories, each of which were significantly different at either the one or five per cent level; fears, school adjustment, sensitivity and excitability, psychosomatic symptoms, and social adjustment.

TABLE II.—COMPARISON OF THE ANONYMOUS GROUP WITH THE SIGNATURES-REQUIRED GROUP: UNFAVORABLE RESPONSES TO THE PERSONALITY QUESTIONNAIRE, FIRST TESTING

Category	Anonymous		Signatures Required		Chi Square of Difference	P	Significance Level
	N	Mean	N	Mean			
Personality Score*	244	8.72	244	9.42	3.62	.20	NS
Cheating Attitudes	244	1.14	244	1.59	2.29	.30	NS

* Includes the following sub-categories: fears, school adjustment, sensitivity and excitability, psychosomatic symptoms, and social adjustment.

TABLE III.—COMPARISON OF SECTIONS OF THE INDIRECT AND DIRECT FORMS OF THE PERSONALITY QUESTIONNAIRE THROUGH CORRELATIONS WITH EXTERNAL CRITERIA: FIRST TESTING

Questionnaire Section	External Criterion	N	Tetrachoric r		Chi Square of Difference	P	Significance Level
			Direct	Indirect			
Social Adjustment Problems	Sociometric Rating	488	+.20	−.21	39.52	.01	1 per cent
Cheating Attitude Admissions	Cheating Behavior	403	+.08	+.25	1.03	.90	NS
School Adjustment Problems	Intelligence Quotients	488	+.17	−.19	49.07	.01	1 per cent

TABLE IV.—CORRELATIONS OF TOTAL PERSONALITY SCORES* OF THE DIRECT AND INDIRECT FORMS OF THE QUESTIONNAIRE, RESPECTIVELY, WITH EXTERNAL CRITERIA: FIRST TESTING

External Criterion	N	Tetrachoric r		Chi Square of Difference	P	Significance Level
		Direct	Indirect			
Intelligence Quotients	488	−.15	+.31	69.05	.01	1 per cent
Socio-Economic Status	488	−.13	+.06	69.77	.01	1 per cent
Reading Comprehension	488	−.04	+.28	53.07	.01	1 per cent

* 'Total personality score' includes fears, school adjustment, sensitivity and excitability, psychosomatic symptoms, and social adjustment.

rate other children, as instructed, and that the superior children in many cases do so rather tolerantly. One finds it difficult to believe, for example, that out of two hundred and forty-four children responding to the indirect form, approximately one-half or more (1) find it difficult to speak out in class or do school work

Table V.—Number of Unfavorable Responses to the Personality and Cheating Attitude Items of the Questionnaire: First Testing*

Nature of Item	Direct N = 244	Indirect N = 244
Finds it hard to speak out in class	85	140
Would rather play by self than with others	18	40
Often becomes gloomy or sad	49	92
Is afraid of high places	58	126
Often has headaches	50	148
Dislikes school	48	76
Cheats to get good marks and please parents	69	32
Often has horrible dreams while sleeping	31	113
Cheats to get good marks and keep up with friends or brothers or sisters	48	33
Cheats if knows that other children are cheating	62	32
Feels that teachers treat him badly or unfairly	14	61
Feels that people often find fault with him	55	53
Is afraid of being followed when walking on the street	43	81
Is afraid of fire	80	148
Often feels ill or weak	26	135
Often day-dreams	87	59
Cheats if believes will not be caught	51	20
Is often bothered with pains in some part of his body	45	147
Feels that nobody really loves him	17	34
Cheats on tests in school if the tests are hard and unfair	84	34
Cheats to get good marks and please the teachers	45	25
Feelings are easily hurt	100	61
Finds it hard to become a leader	77	107
Is afraid of the dark	33	69
Often feels tired when he wakes up in the morning	105	116
Stutters or stammers	14	103

TABLE V.—*Continued*

Nature of Item	Direct N = 244	Indirect N = 244
Has habit of twitching his neck or face	13	46
Bites his nails	85	53
Finds it hard to get along in school	25	57
Cheats if thinks the teacher is unfair	72	30
Often becomes excited	125	120
Finds it hard to make friends	20	81
Other children refuse to play with him	17	70
Often cries	29	67
Finds it hard to do his school work	47	110
Is afraid during bad storms	24	96
Often feels tired during the daytime	33	106

* 'Ringer' items are omitted.

right, (2) often have headaches, pains in some part of the body, or feel ill or weak, (3) often have horrible dreams while sleeping, wake up tired in the morning, and remain so all day, (4) have their feelings easily hurt, and (5) are afraid of high places and of fire.

One may also doubt that in a group of two hundred and forty-four 'normal' fifth-, sixth-, and seventh-grade children (1) more than one-third are afraid of being followed on the street, (2) more than one-third stutter or stammer, (3) other children refuse to play with almost one-third of the group, (4) more than one-fifth bite their nails, and (5) almost one-fifth habitually twitch their necks or faces.

Table V indicates that the approximate two-to-one ratio between number of alleged personality problem 'admissions' on the indirect and direct forms, respectively, was reversed in the case of cheating attitude questions. The reasons for this reversal are not clear from the available data, but it is hypothesized that many of the children in the indirect group did make the identification and transfer when confronted with threatening items, and, realizing the implications involved in the cheating questions, cautiously made fewer admissions than did the direct group. It seems probable that the indirect method will yield some 'neurotic' admissions that the direct method, with its bludgeon-like attack, would miss. But the indirect approach employed in this investigation will score a great many 'normal' responses as 'neurotic' or 'unfavorable.' In more concrete terms, a child, for example, who replies that children who are afraid of high places are 'Not at all queer' is not necessarily, or even probably, admitting that he himself is afraid of high places.

TABLE VI.—NUMBER OF UNFAVORABLE RESPONSES TO THE 'RINGER' ITEMS OF THE PERSONALITY QUESTIONNAIRE: FIRST TESTING

Nature of Item	Direct N = 244	Indirect N = 244
Keeps other children quiet when the teacher is out of the room	99	192
Picks up broken glass lying in the street	72	156
Smiles when things go wrong	92	112
Reports to the police the license numbers of speeding automobiles	27	162
Reports to the teacher other children whom he sees cheating	33	69

Subject to further research involving the effects on the results of an examiner who is a stranger to the children and/or the influences of the individual classroom teachers, Table II points to the conclusion that anonymity of responses is not a major factor in personality and attitude testing at the grade levels here considered. In fact, differences, while not significant by chi-square test, indicate a greater number of unfavorable admissions for the signatures-required group for both the personality and cheating attitude items.

SUMMARY AND CONCLUSIONS

This study evaluates certain aspects of the validity of a direct and of a parallel indirect form of personality questionnaire, most of the items for which originated with eight popular commercial personality questionnaires. The direct form openly asks the child about himself, while the indirect form requests his opinion of other children. The theory advanced by several experimenters but not confirmed in this study, is that children, although ostensibly rating others, will actually reveal their own characteristics. Finally, additional evidence is offered regarding the controversial issue of the effects on responses to personality questionnaires of the requirement or non-requirement of signatures.

Four hundred and eighty-eight fifth-, sixth-, and seventh-grade California public school children were matched in groups of two hundred and forty-four each in order to compare the direct form of the questionnaire with the indirect and the requirement or nonrequirement of signatures. Questionnaire scores for each of the matched groups were compared directly and were correlated with sociometric ratings, intelligence, reading comprehension, socio-economic status, and cheating behavior as measured experimentally.

Correlations of the social adjustment, school adjustment, and cheating attitude sections of both the direct and the indirect forms of the questionnaire with sociometric ratings, intelligence and cheating behavior, respectively, and of total questionnaire scores with intelligence, reading comprehension, and socio-economic status indicate (by these criteria) virtually zero validity for the direct form and a very low degree of validity for the indirect. These correlations, when considered with the item analysis and the significantly different mean scores for the two forms strongly suggest—if they do not conclusively prove—that a number of children sufficient to invalidate the indirect form of the instrument respond as instructed with their opinion of others and generally do not make the personal identification suggested by several previous studies. The blunt, direct form of the questionnaire is regarded as invalid and unpromising and the indirect form as requiring considerable refinement before it can be applied uncritically. Many individual questionnaires which might be useful to a skilled clinician were undoubtedly to be found among both the direct and the indirect groups in this investigation. But the writer joins many previous investigators in cautioning against the making of unwarranted assumptions regarding the validity of this type of instrument for survey purposes.

The mean number of 'neurotic' or unfavorable responses by the signatures-required group exceeded that of the anonymous group, but differences were not significant. Subject to further research involving the effects on responses of an experimenter who is a stranger to the children, and of the classroom teachers concerned, it is concluded that anonymity of response is a minor factor in personality questionnaire testing at these grade levels.

BIBLIOGRAPHY

1) A. Ellis. *A Comparison of the Use of Direct and Indirect Phrasing in Personality Questionnaires*, Psychological Monographs Number 284. Washington, D. C.; The American Psychological Association, Inc., 1947.

2) E. A. Flotow. "Charting Social Relationships of School Children," *The Elementary School Journal*, 46:498–504, May, 1946.

3) H. Hartshorne and May M. Hartshorne. *Studies in Deceit*. New York: Macmillan, 1938.

4) J. L. Moreno. *Who Shall Survive?* Nervous and Mental Disease Monograph, Series 58, Washington, D.C.: Nervous and Mental Disease Publishing Company, 1934.

5) L. M. Terman *et al. Genetic Studies of Genius*, Volume I. Palo Alto: Stanford University Press, 1925.

6) M. J. Van Wagenen. *Unit Scales of Attainment in Reading Comprehension*. Minneapolis: Educational Test Bureau, Inc., 1934.

ACADEMIC ACHIEVEMENTS OF VETERANS AND NON-VETERANS AT THE CITY COLLEGE OF NEW YORK*

LOUIS LAURO AND JAMES D. PERRY
Division of Testing and Guidance
The City College of New York

INTRODUCTION

The question of Federal aid to education and its extension to higher education is still an unresolved issue. It seems appropriate, therefore, to gather as much information as possible on the most recent large-scale subsidization of education through the "G. I. Bill."

The consensus of college professors seems to be that the veterans are better students than their non-veteran classmates. This assumption is supported by published reports, but in many of these studies certain essential controls needed to make this conclusion a valid one have not been introduced.

Unless a non-veteran group is introduced, veterans' scholastic performance can be compared only with their pre-service records, and observed differences might be attributable to factors other than the service experience. Studies which did not introduce a non-veteran control group are of limited interpretative value.[1,8] If a non-veteran group is to be introduced, the factor of scholastic aptitude must also be controlled. This was done by one investigator through the use of analysis of co-variance techniques,[12] by others through matched control groups,[2,5] while still others merely reported mean scholastic aptitude test scores or high-school averages for the groups under consideration.[6,15,16] These investigators reported veterans to be equal to, or superior to, non-veterans. One report, especially noteworthy because it comprised several thousand students in colleges throughout the country, indicated that differences between veterans and control groups of non-veterans were consistently small and in favor of the veterans in the majority of cases.[12] In a special program for veterans with irregular scholastic records, the veterans achieved a higher mean college average than did women in regular attendance at the

---

* The writers wish to thank Dr. Louis Long for his many constructive suggestions, and the staff members of the Veterans Counseling Office for their assistance in recording the data upon which this study is based.

college, in spite of the fact that the mean high-school average of the women was above that of the veterans.[15]

The failure of some investigators[4,7,9,10] to control the factor of scholastic aptitude leaves as an unknown a variable which might affect results. These investigators did find slight differences in favor of veterans, but only one[9] considered the statistical significance of the difference and in this case the difference was not significant. Similarly, in a study which made reference to scholastic aptitude measures for only part of the cases, a very slight difference in favor of veteran students was reported, but a test of significance was lacking.[14]

On the assumption that students' performances in certain college courses are indicative of over-all academic achievements, one course might be selected as a focal point for investigation. However, failure to substantiate the assumption severely limits the interpretation of the results. For this reason, it appears unjustified to extend to over-all achievement the finding that veterans surpassed non-veterans in Freshman English.[13] An investigation centering about grades in general chemistry must be considered inconclusive since there were only eight subjects in the non-veteran group.[3]

An attempt to demonstrate that age was the only factor making for the slight superiority of veterans indicated that non-veterans made better grades than did veterans of the same age.[11] However, when one considers the upward trend of grades through the later college years, it becomes apparent that the failure to take grade placement into account biases the results against veterans. Furthermore, other investigators found no correlation between age and academic achievement of veterans[5] or slight negative correlations for both veteran and non-veteran groups.[2]

There is a good deal of variation in the relative scholastic achievement of veterans and non-veterans from college to college. No institution reported any very dramatic results one way or the other; however, slight differences in favor of the veterans seem to be the norm. Age does not appear to be a factor in causing these differences. Apparently the service experience, *per se*, does contribute to them.

SCOPE OF THE STUDY

The purposes of the present study are: (1) to investigate the average grade-point differences between pre- and post-service periods of veterans attending City College (2) to compare veterans' averages with non-veterans' averages, and (3) to determine whether there are any differences between veterans and non-veterans with respect to choice of course of study (including nature of changes in study program).

The subjects of this investigation are students who were enrolled as lower seniors in the fall semester of 1948 in one of the following courses of study: science, social science, or technology. The records of six hundred twenty-nine male students were obtained. However, a number of these students were not included since they fell into one or more of the following categories:

1) Students who did not attend City College before entering service. They had either entered college for the first time after their return from service or had taken a full semester's work or more at another institution.

2) Students who had attended Evening Session at City College before entering the service or who had less than ten pre-service credits.

3) Students who were taking training under Public Law 16 (the law covering training of disabled veterans).

4) Students who were twenty-one years of age or more at time of entrance.

Data relating to the final sample will be found in Table 1.

TABLE 1.—DESCRIPTION OF FINAL SAMPLE

Course	Classification of Students	Original Sample	Number Excluded	Final Sample
Science	Veterans	66	18	48
Science	Non-Veterans	44	2	42
Social Science	Veterans	104	45	59
Social Science	Non-Veterans	85	13	72
Technology	Veterans	240	120	120
Technology	Non-Veterans	90	17	73

Most of the veterans entered the College between September, 1941, and September, 1944, while most of the non-veterans entered between February, 1945, and February, 1946. The mean age at entrance for the two groups was approximately seventeen years and four months, with standard deviations varying from seven to ten months. High-school averages and scores on the ACE Psychological Examination were recorded for all those students for whom this information was available. In Table 2, the mean high-school average and the mean score on the ACE Psychological Examination are listed for veterans and non-veterans.

TABLE 2.—MEAN HIGH-SCHOOL AVERAGE AND MEAN SCORE ON ACE PSYCHOLOGICAL EXAMINATION FOR VETERANS AND NON-VETERANS

Course	Measure of Scholastic Aptitude	Veterans			Non-Veterans				
		N	M	SD	N	M	SD	$\frac{d}{\sigma d}$	P
Science	ACE	40	121.8	18.4	39	124.7	21.7	.64	.26
	High-school Average	38	83.7	4.0	40	85.1	4.5	1.46	.07
Social Science	ACE	55	116.8	16.8	70	117.0	22.1	.06	.48
	High-school Average	50	82.8	3.9	68	83.4	4.8	.75	.23
Technology	ACE	97	119.7	16.7	65	126.0	15.9	2.42	.01
	High-school Average	94	83.8	3.8	63	84.9	4.1	1.69	.05

The differences between the means and the reliabilities of these differences are also presented in Table 2. The greatest difference between veterans and non-veterans is found in the technology group, where significant differences between the mean high-school averages and mean scores on the Psychological Examination are found in favor of the non-veterans. These differences may be due to variations in admission standards.

METHODS OF TREATING THE DATA

The effect of variations in amounts of college work, taken in the pre- and post-service periods, upon academic achievement was partially controlled by dividing the veterans into two groups. Group A consisted of those veterans who had completed eleven to thirty-two credits in the pre-service period. Group B consisted of those who had completed more than thirty-two credits in the pre-service period. In order to compare Groups A and B in terms of scholastic ability, the mean of the scores on the Psychological Examination was computed for each group, according to course of study undertaken. The difference between the means of Groups A and B was 0.1 for the science students, 4.1 for the social science students, and 1.6 for the technology students. None of these differences were found to be statistically significant.

The pre-service grade-point average and the post-service grade-point average for veterans were determined by dividing the grade-point* by the number of credits taken during the appropriate period.

In order to study the question of whether college grades tend to improve as students progress in their college work, the college grade-point average of the non-veterans was determined at the end of the first, second and third years. The procedure used in calculating these grade-point averages was the same as that for the veterans.

RESULTS

A trend among veterans toward professional and technical training, found by Tibbets and Hunter,[14] is also apparent among the veterans in this study. Of the three courses of study which students in this investigation were taking, technology was being taken by more than a proportionate number of veterans, while the non-veterans had tended to take science and social science courses. This trend held despite the fact that many veterans had transferred from the School of Technology to other courses of study. Table 3 indicates the number that have transferred to science or social science in their most recent degree change. The reason that the number of veteran technology students still remains high, despite this drainage, is that the vast majority of them entered college during the war years, when interest in engineering was at its peak. The Registrar's tabulation of degree choices of entering freshman for the semesters beginning February, 1942, to February, 1944, shows that forty-four to forty-

---

* The grade-point was computed by giving each credit of *A* a weight of plus two, each credit of *B* a weight of plus one, each credit of *C* a weight of zero, each credit of *D* a weight of minus one, and each credit of *F* a weight of minus two; the sum of the weighted credits is the grade-point.

TABLE 3.—NATURE OF MOST RECENT CHANGE OF COURSE BY VETERANS AND NON-VETERANS

Course		Veterans		Non-Veterans	
Changed to:	From:	N	Per Cent of Sample Transferring	N	Per Cent of Sample Transferring
Science	Technology	26	54	8	19
	Arts	1	2	0	0
	Business Administration	0	0	1	2
		27	56	9	21
Social Science	Technology	11	19	2	3
	Science	19	32	8	11
	Arts	6	10	5	7
	Business Administration	3	5	6	8
	Education	2	3	4	6
	Unspecified	1	2	0	0
		42	71	25	35
Technology	Science	7	6	0	0
	Unspecified	2	2	0	0
		9	8	0	0

nine per cent of all freshmen entered the School of Technology, whereas in the semesters beginning February, 1945, to February, 1946, inclusive, only twenty-eight to thirty-four per cent of all freshmen entered this School. These two periods correspond closely to the years when the veterans entered and to those when the non-veterans entered, respectively. The veterans were apparently not influenced by the post-war trend away from technology. Table 3 shows that very few students transferred to the School of Technology, so that almost all of the students in this School at the time of this study began their college careers in it.

In Table 4, the mean pre-service grade-point average of the veterans in the School of Technology is presented along with the mean grade-point average of non-veterans in the same School for two equivalent periods of time. Differences of .21 (for Group A) and .22 (for Group B) of a letter grade are found in favor of the non-veterans; both differences are significant at the one per cent level. In the post-service period, however, veterans in Group A exceeded the scholastic achievement of non-veterans by .13 of a letter grade (P = .03). On the other hand, veterans in Group B did about as well in the post-service period as non-veterans in an equivalent period. The mean high-school average and ACE Psychological Examination score of veterans and non-veterans in the School of Technology (see Table 2) presaged poorer scholastic achievement on the part of veterans. Despite a slightly lower mean 'scholastic aptitude,' veterans equalled or exceeded the achievement of non-veterans.

The data are not as clear-cut in the case of the other degree groups (Table 3). Fully seventy-one per cent of the veterans and thirty-five per cent of the non-veterans taking the social science course had changed to it from another course; fifty-six per cent of the veterans and twenty-one per cent of the non-veterans taking the science course had changed to it from another course. This means that there has been a wide range of subjects taken in the first year or two; the difference is especially marked between those who had taken an engineering course and those who had taken a non-engineering course. These differences must be borne in mind even when comparing veterans with non-veterans taking the same course.

The veterans who were taking a science or social science course as of September, 1948, and who had taken a year or less of college

TABLE 4.—MEAN PRE- AND POST-SERVICE AVERAGES FOR VETERANS, MEAN AVERAGES FOR NON-VETERANS FOR EQUIVALENT PERIODS, AND TOTAL AVERAGES FOR VETERANS AND NON-VETERANS

	Group		N	Pre-Service Periods and Equivalents				Post-Service Periods and Equivalents				Total Period			
				Mean No. of Credits	SD	Mean Grade-Point Average	SD	Mean No. of Credits	SD	Mean Grade-Point Average	SD	Mean No. of Credits	SD	Mean Grade-Point Average	SD
Veterans	BS	(A)[1]	26	22	7	−.12	.73	67	6	.42	.30	90	5	.30	.28
Non-Veterans[2]	BS		42	33	6	.68	.59	65	7	.74	.59	98	5	.71	.56
Veterans	BS	(B)[1]	22	48	14	.35	.41	46	13	.56	.66	94	5	.48	.47
Non-Veterans[3]	BS		42	67	10	.66	.58	31	10	.80	.66	98	5	.71	.56
Veterans	BS	(A&B)[1]	48	34	17	.10	.65	57	15	.48	.50	92	6	.38	.39
Veterans	BSS	(A)[1]	28	20	7	−.08	.78	71	7	.65	.44	91	4	.46	.45
Non-Veterans[2]	BSS		72	30	5	.29	.50	70	7	.61	.37	100	5	.53	.35
Veterans	BSS	(B)[1]	31	48	10	.20	.52	47	12	.65	.60	95	6	.45	.47
Non-Veterans[3]	BSS		72	62	9	.36	.41	38	11	.75	.41	100	5	.53	.35
Veterans	BSS	(A&B)[1]	59	35	17	.07	.67	58	16	.65	.53	93	6	.46	.46
Veterans	Tech.	(A)[1]	76	21	7	.21	.69	69	7	.50	.44	90	7	.44	.44
Non-Veterans[2]	Tech.		73	32	5	.43	.58	71	8	.37	.43	104	6	.40	.42
Veterans	Tech.	(B)[1]	44	46	9	.23	.45	49	11	.32	.43	94	9	.25	.29
Non-Veterans[3]	Tech.		73	65	7	.45	.42	38	10	.29	.51	104	6	.40	.42
Veterans	Tech.	(A&B)[1]	120	30	14	.22	.61	62	13	.43	.45	91	8	.37	.40

[1] Group A is composed of veterans who had completed eleven to thirty-two pre-service credits; Group B, of those who had completed thirty-three or more pre-service credits.

[2] The equivalent of a pre-service period was obtained by determining the average for the first year; the equivalent of a post-service period was obtained by determining the average for the second and third years.

[3] The equivalent of a pre-service period was obtained by determining the average for the first two years; the equivalent of a post-service period was obtained by determining the average for the third year.

work in the pre-service period, had mean grade-point averages for this period of −.12 and −.08, respectively, (equivalent to a letter grade of slightly less than 'C'). Non-veterans taking a science course obtained a mean grade-point average of .66 (B−) for an equivalent period, and those taking a social science course, a mean grade-point average of .29 (C+). In the case of the veteran group, the uncertainty of whether or not they would be drafted before the end of the semester, coupled with the enrollment of many of them in the Engineering School, when a large percentage probably lacked the aptitudes for technological training, are suggested factors contributing to poor scholarship. On the other hand, those of the veteran group who were able to complete more than one year of school work in the pre-service period fared much better during this time with mean grade-point averages of .35 for the science and .20 for the social science groups. However, in an equivalent period, the non-veterans still surpassed the veterans with mean grade-point averages of .66 for the science and .36 for the social science groups.

The post-service period gives an entirely different picture. Groups A and B of veterans taking a science or a social science course all showed increases in mean grade-point averages over the pre-service period. The non-veterans taking a science course were apparently a superior group from the beginning. Although both Groups A and B of the veterans taking a science course showed a marked improvement in the post-service period, they were not able to equal the consistently high performance of non-veterans. More than half of the veterans taking a science course had transferred from the School of Technology; these had started off with a poor adjustment to school work. This poor initial performance may have projected its effects to the time after the transfers had been made. Of veteran students taking a social science course, Group A obtained a mean grade-point average of .65 in the post-service period while non-veterans taking the same course obtained a mean grade-point average of .61. However, this slight difference in favor of the veterans is not statistically significant ($P > .05$). Group B also obtained a mean grade-point average of .65 in the post-service period but non-veterans obtained a mean of .75; the difference between the means is significant ($P = .01$).

Group A of the veterans taking a technology course surpassed the B group in mean post-service grade-point average, although the mean pre-service grade-point averages of both groups had been very close. Group A, it may be recalled, is the group which had one year or less of college training before entering service. Therefore, for the technology students, the service experience appears to have had a beneficial effect on academic achievement. The comparison of Groups A and B of veterans taking the other courses is complicated by the variation in pre-service mean grade-point averages. Group A of veterans taking a science course showed a greater improvement in the post-service period over the pre-service period than did Group B. However, Group B still surpassed Group A in mean grade-point average. On the other hand, Group A of the veterans taking a social science course, whose academic achievement was less than that of Group B in the pre-service period, equalled Group B in the post-service period. The difference with respect to relative achievements of Groups A and B between veterans taking a science course and those taking a social science course may be explainable in terms of differences between the nature of the two courses. Success in advanced work in the science course is more dependent upon a good foundation in certain elementary courses than is the case in the social science course. In the social science course, the materials covered in the various subjects are not as interdependent.

SUMMARY AND CONCLUSIONS

The interpretation of many studies dealing with the relative scholastic achievement of veterans and non-veterans is limited because such factors as age and scholastic aptitude have not been adequately controlled or because tests of the significance of obtained differences have not been made. In general, it appears that the academic achievements of veterans were slightly superior to those of non-veterans.[2,6,12,15] Although it was suggested that age alone was responsible, this thesis was not supported in fact. The service experience itself seems to have contributed to the better performance of veterans. The wide range of experiences, travelling, and constantly making adjustments, probably had their effect in enabling the veteran student to derive more from his education than the non-veteran.

The present study includes four hundred fourteen male students taking science, social science, or technology courses. The preference of veterans for technological subjects, reported by Tibbets and Hunter,[14] shows up in the present study also. This trend seems to be a carry-over from the war-time tendency toward these fields.

In all instances, the mean post-service grade-point average for veterans was higher than their mean pre-service grade-point average. Furthermore, the veterans taking courses in technology exceeded their non-veteran classmates despite a lower mean high-school average, and a lower mean score on the Psychological Examination. However, veterans taking a science or social science course were not able to equal the consistently high performance of non-veterans. Factors that may contribute to this difference between students taking a technology course and those taking either a science or social science course are discussed.

REFERENCES

1) ———. "Academic Achievements of Veterans at Cornell University." *School and Society*, LXV (1947), 101–102.

2) Edward L. Clark. "The Veteran as a College Freshman." *School and Society*, LXVI (1947), 205–207.

3) Paul E. Clark, and Bernard A. Staskiewicz. "Achievements of Veterans in General Chemistry." *School and Society*, LXV (1947), 482–484.

4) Stephen E. Epler. "Do Veterans Make Better Grades than Non-Veterans?" *School and Society*, LXVI (1947), 270.

5) Norman Garmezy and Jean M. Crose. "A Comparison of the Academic Achievement of Matched Groups of Veteran and Non-Veteran Freshmen at the University of Iowa." *Journal of Educational Research*, XLI (1948), 547–550.

6) Arthur M. Gowan, "Characteristics of Freshmen Veterans." *Journal of Higher Education*, XX (1949), 205–206.

7) M. G. Orr. "Grade-point Averages of Veterans at Oklahoma Agricultural and Mechanical College." *School and Society*, LXVI (1947), 94.

8) Frederick D. Pultz. "Veterans in the Ohio State University College of Education." *Educational Research Bulletin*, XXVI (1947), 153–156.

9) G. W. Read. "Scholastic Achievement of Veterans." (Abstract) *American Psychologist*, I (1946), 452.

10) Svend Riemer. "Married Veterans are Good Students." *Marriage and Family Living*, IX (1947), 11–12.

11) Robert H. Schaffer. "A Note on the Alleged Scholastic Superiority of Veterans." *School and Society*, LXVII (1948), 205.

12) William B. Schrader and Norman Frederiksen. "The Comparative Achievement of Veteran and Non-Veteran Students in College." (Abstract) *American Psychologist*, III (1948), 259.

13) Edgar A. Taylor. "How Well Are Veterans Doing?" *School and Society*, LXV (1947), 210–211.

14) Clark Tibbets and Woodrow W. Hunter. "Veterans and Non-Veterans at the University of Michigan." *School and Society*, LXV (1947), 347–350.

15) Ruth G. Weintraub and Ruth E. Salley. "Hunter College Reports on Its Veterans." *School and Society*, LXVIII (1948), 59–63.

16) Ernest L. Welborn. "The Scholarship of Veterans Attending a Teachers College." *Journal of Educational Research*, XL (1946), 209–214.

READABILITY OF CHILDREN'S TEXTBOOKS*

EDMUND W. J. FAISON
George Washington University

The purpose of the present study is to compare the readability of the texts currently used in the fifth, sixth, seventh, and eighth grades of two school systems by application of the Flesch readability formulas.

In 1948, Flesch[3] reported a revision of his readability formula. This readability yardstick has two indices: 1) the measure of 'Reading Ease,' and 2) the measure of 'Human Interest' (hereafter referred to as RE and HI, respectively). The RE formula involves the average sentence length in words and the average word length in sentences. The elements of the HI formula are the average percentages of 'personal words' and the average percentage of 'personal sentences.' Flesch developed this formula only after careful analysis of all of the previous attempts to measure readability, and it is by far the most objective and the easiest to apply of all readability measures.

The application of this formula to children's textbooks is appropriate. The scale offers educators a means of objectively placing texts on a scale of comprehensibility; and it will also point up to the authors of these texts the parts which need special attention when they are writing for a particular level. With this scale as a criterion the proper grade-placement of certain books and of certain types of material can be determined. In this respect, Ayer[1] found that, due to use of figurative language and abstract words and concepts, pupils with normal seventh-grade reading ability could answer only thirty-one per cent of the questions based on original fifth-grade histories used throughout the country.

SAMPLING AND PROCEDURE

The books of two city school systems in the vicinity of Washington, D.C. were used in this study. In the schools of City A there are no assigned books for each grade. The teacher is permitted to select whatever books he wants to use. Therefore, every class may use different texts. A teacher may use one basic text and have several other texts for supplemental reading. This made the problem of sampling difficult. The books selected were the ones which the bookroom clerk estimated were called for more often than the others. In some cases, this made it possible to use the books of a specific series in this study.

---

* This is an abstract of an M.A. thesis done at George Washington University. The author gratefully acknowledges the assistance of Dr. E. Lakin Phillips who helped direct the study.

In City B, standard texts were used in every grade. The texts used in this study were obtained from one elementary school for the fifth- and sixth-grade books, and from a junior high school for the seventh- and eighth-grade books in both cities.

The fifth grade was selected as the lowest grade for the study, inasmuch as Flesch devised his formulas with the top value of 100 representing the score of books suitable for one who has completed the fourth grade (one who is barely functionally literate in terms of the United States Department of Census).

The sampling of the sixth grade was made in order to ascertain what changes in reading ease and human interest take place in books used in the elementary schools. Books used in the seventh grade were sampled in order to learn what changes take place in reading ease and human interest in the jump from elementary schools to junior high schools; and the eighth grade, in order to determine if a greater difference in human interest and reading ease takes place in junior high school than it does in elementary schools.

PROCEDURE IN RATING BOOKS

The formulas for rating RE and HI are as follows: $\text{RE} = 206.835 - .846\ \text{wl} - 1.015\ \text{sl}$ (where *wl* equals the average word length in syllables per 100 words and *sl* equals the average sentence length in words). $\text{HI} = 3.365\text{pw} + .314\text{ps}$ (where *pw* equals the average per cent of personal words and *ps* equals the average per cent of personal sentences).

A score of 100 with the RE formula corresponds to the prediction that a child who has completed the fourth grade will be able to answer correctly three-fourths of the test questions to be asked about the passage that is being rated. An HI score of 100 indicates that the passage has enough human interest to suit the reading habits and skills of one who has completed the fourth grade.[3]

A total of thirty-eight books were rated. One book was selected for each school subject from each grade for each school system. Thirty samples of one hundred words each were selected on a strictly numerical basis from each book. Each sample started at the beginning of a paragraph. The number of syllables, number of sentences, number of personal words, and the number of personal sentences per each one hundred words were counted.

The exact procedure recommended by Flesch was used in all cases with the exception of the mathematics texts. It was considered advisable to apply a special procedure to them so that a comparison could be made with the other texts.

A large portion of all mathematical books consisted of addition or subtraction problems such as:

$$\begin{array}{r} 2735 \\ 4624 \\ 3552 \\ +6189 \\ \hline \end{array}$$

It was felt that these numbers had to be considered and could not be skipped over as Flesch recommends. Also, if each horizontal row was considered as one four-place number, a large number of syllables (and consequent increase in difficulty) would occur which apparently is not the case. In this study, an addition or subtraction problem which had to be completed by the reader was computed on the following basis: Every individual number instead of every horizontal row was counted as a word. This procedure was followed because every student reading these problems would read each vertical column one number at a time, as he actually performs the addition. Each problem was counted as a sentence since it represented a complete unit. This is a somewhat arbitrary procedure, but it appeared to be the most defensible one under the circumstances.

In counting personal words all personal pronouns and nouns of a masculine or feminine gender were included. The group words 'folks' and 'people' (with a plural verb) were also counted. Personal sentences were those which were spoken as indicated by quotation marks or otherwise; questions, commands, requests, etc. directly addressed to the reader; exclamations; and grammatically incomplete sentences whose full meaning had to be inferred from the context.

RESULTS

In comparing school systems, (Table I) the texts of the two school systems revealed few large differences in ease of reading. The books from both systems showed a progressive decrease in reading ease from the fifth through the eighth grades, with the smallest difference between the seventh and eighth grades.

TABLE I.—MEAN SCORES OF READING EASE AND HUMAN INTEREST OF ALL TEXTS IN CITY A AND CITY B, GRADES FIVE THROUGH EIGHT

Grade	City A		City B	
	RE Score	HI Score	RE Score	HI Score
5th	78.5	29.8	76.0	31.0
6th	71.5	27.0	74.8	34.0
7th	68.6	39.0	69.8	35.3
8th	68.6	32.3	68.8	33.6
Mean	71.8	32.0	72.7	33.4

The reading ease scores are in the upper bracket of the range Flesch reports as 'fairly easy' in the fifth grade and steadily drop to the upper portion of what Flesch calls the 'standard' range in the seventh and eighth grades.

Examination of the table shows that the HI scores of books of each school system reveal different patterns. These patterns, however, are difficult to analyze. Flesch defines a score of HI 100 as the score which a book must obtain if it is to hold the interest of a person with a fourth-grade education. This has been interpreted to mean that as a person becomes better educated, the reading material can gradually decrease in human interest and still hold the reader's attention.

All of the mean human interest scores for each grade are within the category which Flesch calls 'interesting.' The range of the scores is not great, but the patterns within the range are of interest. In City B, the fifth-grade texts were the least interesting. A gradual increase in human interest occurred up to the eighth grade, where it dropped to a level slightly below that of the sixth grade. Although the differences are small, they show a trend.

In City A, the sixth-grade texts were less interesting than those of the eighth grade. But the textbooks of the junior high school (seventh and eighth grades) were more interesting than those of the elementary school.

With the exception of the seventh-grade score, the schools of City B used more interesting texts on the whole. The high seventh-grade human interest score (City A) was due to the very large increases in HI in science and literature from the sixth to the seventh grade.

ANALYSIS OF RESULTS BY GRADES

Results by grades are given in Table II. Although the average reading ease scores dropped consistently, many discrepancies occurred among the individual subjects. In City B, English became easier to read as the grade increased. This result was produced mainly by the larger number of syllables in the lower books.

TABLE II.—COMPARISON OF MEAN READING EASE AND HUMAN INTEREST SCORES OF THE TEXTS OF FIVE SUBJECTS TAUGHT IN TWO SCHOOL SYSTEMS FROM THE FIFTH TO THE EIGHTH GRADES INCLUSIVELY

Subject	Location	5th		6th		7th		8th	
		RE	HI	RE	HI	RE	HI	RE	HI
English	City A					75	50	70	34
	City B	76	33	77	45	77	52	83	57
History	City A	85	51	71	29	68	34	60	23
	City B	81	29	75	21	68	22	60	18
Math.	City A	71	30	69	26	55	19	58	19
	City B	69	21	78	30	58	34	62	20
Reading or Lit.	City A	79*	29*	74	39	74	55	80	57
	City B	86	47	81	57	74	42	72	43
Science	City A	79	09	72	14	74	37	75	28
	City B	68*	25*	63	17	72	26	67	30

* In City A, Social Studies is taught instead of Reading in the fifth and sixth grades. In City B, Geography is taught rather than Science in the fifth and sixth grades. These subjects are listed for comparison.

In both cities the reading ease of mathematics increased from the seventh to the eighth grades. This increase was due to the change in style of arithmetic problems. Instead of listing columns of problems to be added or subtracted, the eighth grade books more often expressed the problems in a realistic fashion with the result that a little story was told about every problem situation and the effect of the actual numbers was minimized in figuring reading ease.

This result was achieved primarily because of the difference in the subject matter taught. In the seventh grade common fractions, decimal fractions, integers, measurement, and percentages are the primary topics considered. In the eighth grade, banking, insurance, taxes, geometry, investments, and mensuration are taught. It can readily be seen that the topics in the eighth grade cannot have problems which are expressed in rows with instructions to perform the necessary operations.

There are other inconsistencies which cannot be explained on any basis other than the style of the authors. The largest drop in reading ease occurred in mathematics between the sixth and seventh grades. Here the child takes only arithmetic in the elementary school, but when he begins junior high school he is confronted with the much broader and more difficult subject of mathematics which includes all of the topics already mentioned. This increase in difficulty of the subject is reflected in the reading ease score.

It would seem that the factor of human interest has been practically ignored by the writers of all texts with the exception of the subject of mathematics. As is expressed in the preface of many of the mathematics books, there has been a conscious effort to state as many exercises as feasible in problem situations rather than just problems. These authors have realized that real learning takes place more rapidly when the child is projected into the problem situation. It is a real tribute to these authors that they have made their subject matter more interesting than science or history in some cases.

Only the fifth grade science text of City A fell into the 'dull' category. It included few personal sentences or words.

ANALYSIS OF RESULTS BY SCHOOL SUBJECTS

The over-all trends are shown in Table III. It can be seen from this table that the most difficult subject was mathematics.

The average difficulty for all four grades was higher than the average eighth-grade score for all subjects.

TABLE III.—COMPARISON OF THE MEAN READING EASE SCORES AND MEAN HUMAN INTEREST SCORES OF THE TEXTS OF FIVE SUBJECTS FROM THE FIFTH THROUGH THE EIGHTH GRADES OF THE CITY A AND CITY B SCHOOL SYSTEMS

	Math	Science	History	English	Literature
City A					
RE	63.25	75.0	71.0	71.0	76.75
HI	23.25	22.0	34.25	42.0	45.50
City B					
RE	66.75	67.50	71.0	78.25	78.25
HI	26.25	24.50	22.25	49.25	47.25
Mean Score					
RE	65.0	71.25	71.0	74.63	77.50
HI	24.75	23.25	28.25	45.63	46.37

The procedure in rating mathematics books was different from that used for the others. The modification of the method recommended by Flesch was incorporated to increase the reading ease scores so that they could more readily be compared with the scores of the other texts. If the procedure Flesch recommends had been applied to the mathematics books, the scores would have been even lower on the reading ease scale. With either procedure, however, mathematics would still be considered the most difficult subject when using the reading ease score as a criterion.

History, with a reading ease mean score of 71.0, was the second most difficult subject. This score according to Flesch's norms would be on the border between 'standard' and 'fairly easy.'

The history average was the same in both cities, but in City B, history was the third most difficult subject to read, with science preceding it; whereas in City A, history and English were tied for second and third (2.5) place in the rank order scale of difficulty.

Science books had an over-all mean of 71.25. This merits third place on the scale, but the difference of .25 between this score and the mean history score is too small to be considered significant.

In City B, science ranked second, while in City A, it ranked fourth in the order of reading difficulty.

In the fifth and sixth grades in City B, geography is taught in place of science, and in this study geography was considered with the science group.

English was the fourth of the five subjects on the composite mean rating scale. There was, however, a marked difference between the two school systems. English was tied for the second and third positions (actually position 2.5) with history in City A; and tied with literature for the fourth and fifth places (actually 4.5) in City B.

Literature was found to be the easiest subject with an average reading ease score of 77.5. This over-all average is at the mean of all the subjects of the fifth grade.

Literature, as such, is not taught in the elementary schools. In City B, fifth- and sixth-grade readers are used. These readers contain light stories about the activities and accomplishments of children. In City A, the elementary schools have books which are entitled 'Social Studies.' These books are used as readers, but they contain stories which are more directly related to educational topics. It is not surprising, therefore, to find that the City B elementary schools have higher reading ease scores for their readers than in City A.

Mean human interest scores are also given in Table III. The mathematics textbooks had a human interest value of 24.75. On the rank-order scale which begins with the least interesting subject, mathematics texts ranked second with a value 1.5 points higher than science.

In City A, the human interest score of mathematics ranks exactly as it does in the composite mean. In City B, mathematics is third, falling behind history and science.

History ranked third on the composite human interest scale with the score of 28.25. This agreed with the result obtained in City A; but in City B, history was the dullest subject in terms of human interest. This result was caused by the long sentences and few personal references.

Science was the dullest subject in the composite scale in City A. This result was due mainly to the lack of personal words. It is very difficult to be personal about scientific facts unless an experiment is described by the worker who performed it. In City B, science was second with history preceding it by 2.25 points.

In City A, literature was slightly higher than English, and was, therefore, fifth in that series and in the composite scale. But in City B, literature was slightly lower than English, but not enough to reverse the positions on the composite scale. The averages of both subjects fall in the group Flesch calls 'highly interesting.' These two subjects are nearly 20 points higher than the other subjects in human interest.

SUMMARY

All books rated for reading ease fell within the range of Flesch's categories 'fairly difficult' to 'easy.' The norms of each grade for both school systems showed a consistent decrease with the scores gradually leveling out between the seventh and eighth grades. The subjects were ranked in order of difficulty (reading ease score) and the following results were obtained: Mathematics (most difficult), history, science, English, literature (easiest).

The mathematics average for all four grades was lower than the average of all subjects of the eighth grade. The literature average for all grades was approximately that of the average of all of the subjects of the fifth grade.

No definite pattern was shown in human interest scores. All of the averages for the individual grades were in the 'interesting' range, however. The mathematics texts were the only ones in which a conscious attempt had been made by the author to personalize the material presented, but even there no system seemed to stand out.

BIBLIOGRAPHY

1) A. M. Ayer. *Some Difficulties in Elementary School History.* Contributions to Education. No. 212. New York: Bureau of Publications, Teachers College, Columbia University, 1926.

2) J. N. Farr and J. J. Jenkins. "Tables for use with the Flesch readability formulas." *J. Appl. Psychol.*, 1949, 33, 275–278.

3) R. Flesch. "A new readability yardstick." *J. Appl. Psychol.*, 1948, 32, 221–233.

SELECTION OF STUDENTS FOR A TRADE AND INDUSTRIAL EDUCATION CURRICULUM

H. S. BELMAN AND R. N. EVANS
Purdue University

Numerous articles by educators have appeared, suggesting that it would be desirable to select students for teacher training in industrial education, rather than to accept all those who apply. However, few concrete proposals or examples have been presented.

In 1942, Pawelek found that "in eighty-two per cent of the industrial arts departments, any student admitted to the university is automatically admitted to the department." When one hundred leading educators in the industrial arts field were polled, however, more than ninety-five per cent of them favored setting up definite admission standards.[2]

Land has been one of the few leaders in the field of industrial education who has reported on a comprehensive, objective method of selection. He states that prospective teachers of vocational industrial subjects at Pennsylvania State College are given a battery of tests which include "a four-hour trade performance test administered by a competent examiner, a three-hour trade theory examination, a psychological examination, an English test, and an interview by a competent committee of three." This selection procedure is based upon a definite philosophy which might well be adopted elsewhere, for Land goes on to say: "The teacher-training institution which merely conducts courses for the preparation of teachers, without regard to their adequate selection and placement . . . is not meeting its responsibility."[1] However, concrete programs of this or any other type are the exception at present, and even Land makes no reference to studies of the validity and optimum weighting of these predictors.

Traditionally, the selection of candidates for industrial teacher training has been done on the basis of physical appearance, method of shaking hands, and a perusal of assorted grades, tests, and recommendations. Little objective study has been devoted to the relevance of such predictors of success. The fact that a sizeable portion of the men selected in this fashion have been successful is an indication that it has produced some good results, but as competition increases in the field of industrial education, any method of selection which does not yield perfect results should be examined periodically.

STATEMENT OF THE PROBLEM

The Trade and Industrial Education curriculum, Division of Education and Applied Psychology, Purdue University, prepares students for one of three types of employment:

1) Teaching industrial education subjects in the public secondary schools. (a) vocational industrial. (b) industrial arts.
2) Training within industry.
3) School building service (superintendents of buildings and grounds).

Students enter this curriculum from two principal sources: approximately thirty per cent enter the curriculum as freshmen, and approximately seventy per cent transfer from other schools of the University. This study is concerned with the selection of the latter group from among those who apply.

The transfer problem in any large university is a troublesome one. Many students enroll, for example, in engineering, believing that it involves primarily manipulative work. They often persevere in the school of their original choice for three or four semesters, becoming progressively more maladjusted, unhappy, and earning progressively poorer scholastic records. Other students with no particular aptitude for college work also have difficulties with their present programs. Both types are among the applicants for transfer to trade and industrial education. Obviously, it is to the best interest of both the student and the school, that only those with aptitude for teaching in trade and industrial education be selected.

DESIGN OF THE EXPERIMENT

The present research is based upon an unpublished study by H. W. Porter and R. N. Evans[4] in which records of eight tests taken by twenty-seven graduates of the Trade and Industrial

Education curriculum at Purdue were studied. Using the criterion of grade point index after transfer to the Trade and Industrial Education curriculum, the best individual predictor of the eight studied yielded a correlation of .67 with the criterion. This study indicated that students could be selected more effectively than had been possible in the past.

The next problem was to secure more representative data covering a larger number of predictors. With this in mind, arrangements were made to administer a test battery made up of sixteen separate predictors to all World War II veterans and twenty-five per cent of the non-veterans in the T. & I. E. curriculum.* Participation in the testing program was voluntary, but ninety per cent of the eligible students took the entire battery of tests upon which these results are based. Ten additional predictors were obtained from orientation test scores and from student records, making a possible total of twenty-six test scores and other pertinent facts for each of the one hundred seven students who participated.

MEASUREMENT TECHNIQUES AND RESULTS

As in the previous study, the criterion of grade point index after transfer was correlated with each of the predictors, yielding the results shown in Table I. A measure of statistical significance was applied[3] and all of the predictors which proved to be significantly non-zero at the one per cent level were selected for inclusion in a Wherry-Doolittle test selection problem.[5,6] This procedure provides a means by which tests are selected from a battery in the order in which they contribute to the prediction of the criterion, and at the same time eliminates tests which contribute nothing new to this prediction. A further result of this procedure is the determination of the optimum weighting for each test selected.

For students who have been allowed to enroll in Trade and Industrial Education courses on a one-semester trial basis, the prediction sheet shown in Table II is used as a basis for deciding

---

* These tests were administered and scored by the staff of the Veterans' Guidance Center at Purdue. The cost of testing non-veterans is borne by the University.

TABLE I.—STUDYING TRANSFER STUDENTS IN TRADE AND INDUSTRIAL EDUCATION

Index after Transfer = Criterion

Predictors	Correlation with Criterion	*N*
Kuder Mechanical Interest	.0335	107
Kuder Computational	−.0056	107
Kuder Scientific	.0437	107
Kuder Persuasive	.0128	107
Kuder Artistic	.0529	107
Kuder Literary	.0364	107
Kuder Musical	.1309	107
Kuder Social Service	.1670	107
Kuder Clerical	.0059	107
Guilford-Martin O	.0109	107
Guilford-Martin Ag	.1482	107
Guilford-Martin Co	.1659	107
How Supervise	.2893*	107
How I Teach	.3154*	107
Purdue Mechanical Adaptability	.1875	106
Purdue Adaptability	.4120*	102
American Council on Education Psych.	.2631	107
Purdue English	.2995*	99
Purdue Mathematics	.3441*	100
Purdue Physical Science	.2621	82
Index before transfer	.4982*	95
Grade in Introductory Course to Trade and Industrial Education	.5381*	103
Previously dropped from University—yes or no	.2678	107
Failed or passed Engineering Drawing	.4368*	107
Veteran or non-veteran	.0806	107
Definitely chose option—yes or no	.4347*	107

Number for which every predictor was available = 83

* Indicates one per cent level of confidence that r is non-zero.

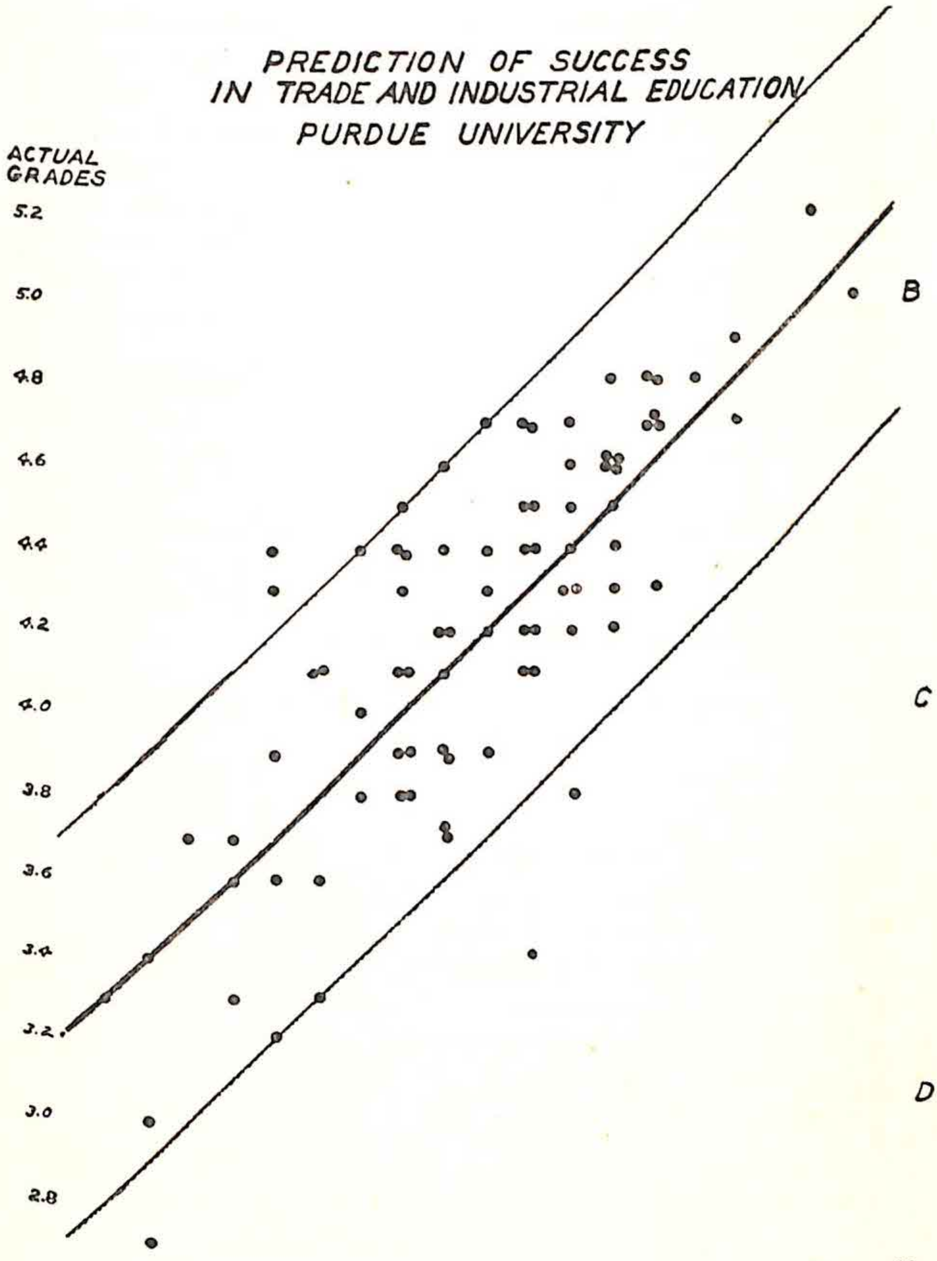
whether or not they should continue. The predictors shown in the table are not adequate for the majority of transfer students, however, since they do not have the opportunity to enroll in Education 30 (Introduction to Trade and Industrial Education) before transferring. For this reason, the statistical computation was done a second time, omitting this predictor, which produced slightly different results. The multiple correlation coefficient for the predictors shown in Table II was .743, while the multiple correlation coefficient for the second method, which did not include the one predictor, was .731.

It should be noted that these correlation coefficients, while far from perfect, are considerably higher than most of those obtained with test batteries now widely used in predicting success in industry, and in colleges, universities, and high schools. However, at least a temporary upper limit has been approached, and it is unlikely that correlations much higher than about .75 will be obtained, until some method of increasing the reliability of grades has been devised and put into effect.

TABLE II.—PREDICTION SHEET FOR USE WITH UNCLASSIFIED STUDENTS

Grade in Education 30 ______ × .2025 = . ______
Index Before Transfer . ______ × .3258 = . ______
Score on *Purdue Adaptability* (per centile, Purdue Seniors) ______ × .0029 = . ______
Score on How Supervise, Form M (per centile, Level II Supervisors) ______ × .0026 = . ______
Score on English Orientation Test (per centile, Purdue Freshmen) ______ × .0026 = . ______
If This Student Has Definitely Chosen an Option, add +.3200 × . ______
Plus 1.4270
Equals Expected Grade Point Average After Transfer . ______

As another way of showing the results obtained, Figure 1 has been prepared. The limit lines have no statistical significance, but are merely plus or minus one-half grade point, and show that if, for example, it is predicted that a student will make a 4.0 grade


PREDICTION OF SUCCESS
IN TRADE AND INDUSTRIAL EDUCATION
PURDUE UNIVERSITY
ACTUAL GRADES
5.2
5.0
4.8
4.6
4.4
4.2
4.0
3.8
3.6
3.4
3.2
3.0
2.8
2.6
B
C
D
3.2
3.4
3.6
3.8
4.0
4.2
4.4
4.6
4.8
5.0
5.2
PREDICTED GRADES

index, it is quite unlikely that he will actually make below 3.5 or above 4.5.

The prediction methods described above are now in use in the Trade and Industrial Education curriculum at Purdue, but the process of further development of these selection techniques is still going on. New transfer students and the remainder of the non-veterans in the school are now taking a different battery of tests. The results achieved will be correlated at a later date with the grade point indices obtained by these students, and the whole selection procedure overhauled in the light of this new data. It is felt that only through a continuing process of experimentation can optimum results be obtained.

CONCLUSIONS

1) The scholastic success of students entering a teacher-training curriculum in industrial education can be predicted with a reasonable degree of accuracy when correct techniques are used.

2) Such predictions can be used with considerable assurance in counselling students; and, when institutional policies permit, these predictions can form a very important part of the student screening and selection procedure.

BIBLIOGRAPHY

1) S. Lewis Land. "The Teacher Training Institution and Postwar Industrial Education," *Industrial Arts and Vocational Education*, 35: 5–7, January, 1946.

2) S. Pawelek. "Some Aspects of Industrial Arts Teacher Preparation." *Industrial Arts and Vocational Education*, 31: 147–9, April, 1942.

3) C. C. Peters and W. R. Van Voorhis. *Statistical Procedures and Their Mathematical Bases*. New York: McGraw-Hill, 1940.

4) H. W. Porter and R. N. Evans. "Predicting Success in the Trade and Industrial Education Curriculum." (Unpublished.) Purdue University, 1949.

5) W. H. Stead, C. L. Shartle, *et al.* *Occupational Counseling Techniques*. New York: American Book Co., 1940.

6) J. Tiffin. *Industrial Psychology*. New York: Prentice-Hall, 1947.

BOOK REVIEWS

MATTHEW N. CHAPPELL. *In the Name of Common Sense: Worry and Its Control.* New York: The Macmillan Company, 1949, pp. 172. $2.75.

The title of this book, *In the Name of Common Sense*, bears little relationship to the actual content; the sub-title, *Worry and Its Control*, indicates more accurately the nature of the material presented.

The author seems not to have made up his mind whether he is writing for professional or lay readers. For example, it is difficult to see what the average person would make out of the following quotation: "Life, Pike states, is a function of four variable factors: (E), the environment; (PC), the physicochemical conditions of the body; (G), the glandular system; and (N), the nervous system. Mathematically, Pike's formula is stated as follows:

$$L = f((E)(PC)(G)(N))$$ "[1]

On the other hand, such passages as the following hardly appeal to the psychologist or the psychiatrist: "The word 'neurotic' is a relic of a mistake made two or three hundred years ago. So are its cousins, 'hysteria,' 'neurasthenia,' and 'psychasthenia.' They are all mistakes of past centuries. Now they are nearly meaningless. They all mean the same thing: that one has developed outstanding skill in making himself emotional and in maintaining that skill. Anyone who tries to make them mean more is talking through his hat."[2]

The book is confusing at many points. Chappell rightly condemns the concept of will power, yet his recommendations for the control of worry sound surprisingly like the admonition to use will power and stop worrying. He urges his reader not to talk about his difficulties, asking him to make a resolution that he will discuss his troubles with no one except his physician. This point of view is not in keeping with modern therapeutic methods. It is rather generally accepted that catharsis is the first step in psychotherapy. One gets the impression from reading Chappell's book that worriers should begin with positive action instead of ending with it. There is little understanding here of the dynamics of anxiety. Insight is referred to but not emphasized.

---

[1] p. 163.

[2] pp. 87–88.

This reviewer sees but little in this book to recommend it. It is too technical for the layman and too superficial and opinionated for the professional reader. Moreover the point of view presented is frequently at variance with well established principles of psychotherapy. HERBERT A. CARROLL

University of New Hampshire

A. ANASTASI AND J. P. FOLEY, JR. *Differential Psychology.* Revised Edition. New York: The Macmillan Co., 1949, pp. 894.

This book is thoroughly revised and greatly enlarged. The revised edition is larger by two hundred seventy-four pages, and each page contains about one-fourth more words than the original edition. Whereas the first edition makes use of five hundred sixty end-of-chapter references, the revised edition presents nearly fifteen hundred. The revised edition contains four additional chapters, the material of which is largely new. These are "Basic Concepts of Psychological Testing," "Biological Factors in Simple Behavior Development," "Psychological Factors in Simple Behavior Development," and "Schooling and Intelligence." Practically all the other chapters are expanded—sometimes by the addition of new sections and sometimes by a complete rewriting.

Style, purpose and general method of treatment of data as found in the original edition are largely preserved. There is the same or perhaps slightly greater emphasis upon methodological problems and interpretation. Certainly both editions amount to source books in their field, but much of their significance lies in the fact that they present the student with a coherent, significant and meaningful account of the data. As such, the revised edition, like the first, should make a useful textbook. It should fill the needs of collateral reading for several courses in psychology. It is a significant contribution. J. B. STROUD

State University of Iowa

JEAN PIAGET AND BÄRBEL INHELDER. *Le Développement des Quantités chez l'Enfant. Conservation et Atomisme.* Paris, 7e, rue de Grenelle: Editions Delachaux & Niestlé S. A., pp. 344.

The specific processes of maturation and of learning through which children's early understandings are formed constitute a subject of great interest to psychologists and teachers. Certainly, all too little is known in this area at present. Piaget and his collaborator contribute here a significant report of research in the formation of quantitative concepts by children, with three principal problems involving experimental induction as the centers of attention.

The first problem attacked by Piaget and Inhelder is this: By what sequential development and at what age may ideas of the permanence of substance, of weight and of volume become understood by children? The technique used was as follows: The child was given a little ball of clay and was directed to make another just like it, "the same size and the same weight." Then one of the balls was altered in shape, either by lengthening it to the form of a sausage, or by flattening it like a pancake, or by some other method. Thereupon, the child was asked whether the two balls still had the same weight, the same quantity of material, and the same volume. He was also urged to explain his answers.

Four important stages were observed in the successive appearance of concepts inherent in this first problem:

1) Before the age of seven or eight, the child understands neither the permanence of the substance, nor of the weight, nor of the volume of the deformed ball of clay.

2) At the approximate age of eight to ten, the child comprehends that the ball is of the same substance, but not that it has the same weight and volume as before.

3) From ten to eleven or twelve, he understands that it is of the same substance and weight, but not yet that it has the same volume.

4) Beginning with age eleven or twelve, the child is able to recognize simultaneously the three forms of permanence of the deformed ball of clay, with a tendency to reduce the idea of substance to those of weight and of volume.

The second aspect of this research also concerns itself with children's ideas regarding the permanence of substance, and, more specifically, with what becomes of sugar when it is dissolved in water. Piaget and Inhelder, working with more than one hundred children from four to twelve, found indications of the same stages in development of understandings as those observed in the experiment with the small balls of clay. At first, the child believes that the substance of the sugar, when dissolved, diminishes to nothing. Later, he understands that the substance is maintained, but thinks that the sugar's weight and volume are lost. Then he adds the idea of the invariability of weight and, finally, of the invariability of volume.

A third area of experimentation reported in this volume is related to children's concepts of the nature of substance. Such materials as a cork, a pebble (smaller but heavier than the cork), and a piece of wood (of a size intermediate between the two, heavier than the cork, but lighter than the pebble) were presented to the children and made the subjects of questioning. Four stages of conceptual growth are recorded, indicating progress from the belief that a body is heavy in proportion to its size, to an understanding of density, dilation and contraction.

The four stages of development which the investigators found in each of these experiments may seem to the reader to be a rather too highly systematized means of reporting the findings. The important conclusions, perhaps, are contained in the observation of a consistent sequential development of the concepts, with a similar pattern in each experiment. Such conclusions may have valuable implications in terms of curriculum planning and the direction of learning activities in the elementary school.

Any experimental work is commendable which adds to our knowledge of how children change through maturation and through perceptual experiences, at what rate and through what sequences they arrive at insightful response to the innumerable problematic situations where they find themselves. Piaget has, in this volume, added notably to his already famous contributions in the study of children's concepts. He and his associate have suggested lines of experimentation which may continue to increase in important ways our understanding of children.

WILLIS N. POTTER
College of the Pacific

SAUL ROSENZWEIG WITH KATE L. KOGAN. *Psychodiagnosis.* New York: Grune and Stratton, 1949, pp. 380.

As a sequel to F. L. Wells' *Mental Tests and Clinical Practice* and Murray's *Explorations in Personality*, this book explains the purpose, materials, obtained data, scoring methods, and interpretation of a representative group of psychological tests as they are employed in up-to-date psychodiagnosis.

This treatment, addressed to beginning students of professional psychology and those in the allied fields of medicine, nursing, social work, etc., is in untechnical language; it excludes standardization procedures and gives little emphasis to theoretical considerations.

A brief discussion of clinical psychology as a diagnostic art is followed by several chapters dealing with tests of general intelligence, measures of intellectual deviation, tests of vocational aptitude and interest, and projective techniques. The process of psychodiagnostic integration is presented in terms of concrete case history examples, illustrating clinical procedure in the synthesis of information obtained from the separate tests and interviews which results in the emergence of a final picture of the personality.

In a retrospective and projected consideration of psychodiagnosis as a science, the author traces the development of the theories of the academic psychologist and the practices of the clinical psychologist as relatively independent movements. A shift in the center of gravity in psychology from emphasis on the generalized average man to the structure and organization of the individual person is recommended as a common effort in which the academic and clinical psychologist can unite for scientific advance.

This explicative but untechnical presentation of psychodiagnostic procedures supplies an important link in the development of a closer liaison between clinical psychologists and professionals in allied fields. The effectiveness of this book as a text book for beginning students in clinical psychology is fortified by the practical presentation of the subject and the interesting and illustrative case material.

The section on vocational testing is a little disappointing. As the author recognizes, the selection of tests is circumscribed and consequently not very representative; furthermore, the reader is left with the impression that vocational recommendations emerge chiefly, if not solely, from the test results. At least a brief mention of client-centered vocational counseling as something more than test administration and interpretation would have been helpful.

William F. Holmes

The Psychological Corporation

Loraine A. Dahl. *Public School Audiometry.* Danville, Ill.: The Interstate Printers and Publishers, 1949, pp. 290.

Up to the present the measurement of hearing status among school children has not received the attention nor the support that it deserves. This is partly due to inconsistent techniques and standards in the various surveys and the programs for conservation of hearing. Nevertheless, it is well established that there are a large number of children of school age who have serious hearing impairments. For purposes of medical care, and for aid in educational and personality adjustment, it is important that these children be identified. The material in the present book should both promote and facilitate more adequate measurement of hearing status.

This book is both a text and a field manual. Part I, on factors to be considered in a hearing conservation program, deals with evidence of the problem, factors hindering the program of hearing conservation, and socio-economic problems of the hard-of-hearing child. The next part considers the techniques for creating interest in the conservation of hearing. Part III presents in complete and clear detail the techniques and principles of audiometry for the discrete frequency audiometer and for group audiometry. The appendixes include material on supplies, testing schedules, and forms of reports. There is an extensive bibliography.

This book presents useful information for all those interested in the hearing problems of school children. No school nurse or other person concerned with audiometry can afford to be without it.

Miles A. Tinker

University of Minnesota

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 42 March, 1951 Number 3

DO ALUMNI AND STUDENTS DIFFER IN THEIR ATTITUDES TOWARD INSTRUCTORS?

A. J. DRUCKER AND H. H. REMMERS*
Purdue University

Among early comments concerning the rating of instructors by students was the criticism that "student judgment may be unripe and different from later judgment. . . "[5] Since the time that statement and others like it were made, the influence of maturity within undergraduate groups upon ratings of instructors, using the Purdue Rating Scale for Instructors, has been investigated. Comparisons of the ratings of freshmen, sophomores, juniors and seniors showed slight differences in ratings of some of the traits, but these were "relatively unimportant as compared with the resemblances among the four classes."[5]

The current investigation was undertaken to test the hypothesis that students' ideas of what constitutes a good instructor change as the students grow older, by comparing student and alumni judgments of the same traits and instructors. The purpose of the Purdue Rating Scale for Instructors is to measure the student's current judgments, or prejudices, as an important aspect of the teacher-student relationship, important as a vital factor in determining the learning process.[6] It has been stipulated that the teacher who uses this scale must constantly bear in mind that the scale is primarily a device for ascertaining student judgment concerning the traits in question. Regardless of what the teacher believes or knows about himself with respect to these traits, the student attitude exists, and exists as a functioning factor in the classroom procedure.[3] Results of this investigation hence cannot be expected to alter the basic purposes

Vol. 42 No. 3 Mar. 51

---

* Dr. N. L. Gage gave major assistance in the design of the experiment.

or aims of the Purdue Rating Scale for Instructors. If, however, it can be shown that students' judgments of the same instructors do not actually change with the years, perhaps more instructors will avail themselves of the instrument with increased confidence and reliance in the judgments of the undergraduates they teach.

To test the above hypothesis, two types of comparisons were to be made: 1) a comparison of the attitudes of contemporary undergraduates toward college instructors with the attitudes of alumni of ten or more years' standing toward the same instructors; 2) a comparison of the desiderata of good college instructors as conceived by contemporary undergraduates and by alumni of ten or more years' standing. By 'attitudes toward instructors' we mean ratings by students and alumni of their instructors on the Purdue Rating Scale for Instructors. By 'Instructors' we mean members of the instructional staff for undergraduates at Purdue University. By 'desiderata of good instructors' we mean ratings as to importance for good instruction of a series of traits of instructors.

PROCEDURE

Since a follow-up study to test our hypothesis was not feasible, it was necessary to obtain the judgments of Purdue University graduates of the years of 1936 to 1939 and compare them with the judgments of current students of the same curricula. In addition it was necessary to use in the study instructors who were on the staff in 1936 as well as now, and who are still teaching the same courses. The procedure involved in obtaining ratings was different in at least two respects from that usually involved in obtaining ratings: 1) one to five ratings were to be made on a single scale, whereas in the past only a single instructor was rated on a scale; 2) the ratings were to be made outside the classroom with the permission of the instructors. Since these differences in procedure were alike for students and alumni making the ratings, they were not expected to have any effect upon any comparisons of ratings by students and alumni. The fact that classroom time was not being used for making the ratings may have had the effect of obtaining ratings on a more representative group of instructors than hitherto has been obtained, since in the past only those instructors who both desired the ratings and were willing to take class time for them have been rated.

Selecting Instructors.—A list was made of all instructors appearing in both the Purdue University Catalogue of 1936 and in the

1948–1949 Catalogue. A letter was then sent to each instructor stating the purpose of the investigation and asking the participation of the instructor whose rôle would be purely a passive one—that of being rated by alumni and students. Of one hundred thirty-nine instructors of whom participation was requested, one hundred two consented to being rated. Of the thirty-seven who did not give consent, several had no classes of undergraduate students for the current year.

Selecting Alumni and Students.—To maximize the number of ratings for the instructors participating in the study, it was decided to select alumni and students from schools and departments in proportion to the representation of eligible instructors. This meant that the students were not selected as strictly representative of the total University student enrollment. It was felt, however, that this was not a major consideration. In a previous study at Purdue University it was found that student ratings were unaffected by a number of investigated factors. Freshmen rated their instructors no higher and no lower than did seniors, male students rated their instructors no differently than did female students, veteran students rated their instructors similarly to non-veterans, and students in the upper half of the class rated their instructors like those in the lower half.[7]

At Brooklyn College, in an extensive teacher-rating project the evidence indicated that the ratings given to an instructor were not affected by the student's college class, sex or by the fact that the course he taught was an elective or prescribed course.[2]

It was decided to send Rating Scales and instructions to five hundred students and to five hundred alumni in the proportions indicated in Table 1 under per cent of eligible instructors. Within schools the sample of students selected was equalized by classes, equal numbers of freshmen, sophomores, juniors and seniors being chosen. For example, in Agriculture, eighteen per cent of the total sample of students or ninety students were to be chosen. This resolved into twenty-two freshmen, twenty-three sophomores, twenty-two juniors and twenty-three seniors from the School of Agriculture. In Engineering, since freshmen are not designated by Engineering department, the entire quota of Engineering students was divided by four to indicate the number of freshmen engineers to select. Then the sophomores, juniors and seniors were divided according to the quotas of their respective engineering schools.

TABLE 1.—COMPOSITION OF THE SAMPLE OF STUDENTS BY SCHOOL MADE TO CONFORM TO THE PER CENT OF ELIGIBLE INSTRUCTORS

School or Department	Per cent Eligible Instructors	Number of Students Selected	Per cent Approximate University Enrollment
Agricultural Engineering	2	8*	1
Chemical Engineering	3	15	8
Civil Engineering	9	34	6
Electrical Engineering	6	22	13
Air Transportation, Mechanical and Aeronautical Engineering	6	22	24
Freshman Engineering		34	
Agriculture	18	90	10
Forestry	2	10	2
Home Economics	6	30	9
Pharmacy	4	20	4
Physical Education	4	12	3
Science	38	198	18
Trade & Industrial Education	1	5	2
Totals	100	500	100

* Numbers in departmentalized engineering groups are one-fourth less because of the allowance for freshman engineers.

Students were selected for participation at random with the help of the current student directory. First the elements of the student sample were listed according to class and curriculum. Then for a given school and a given class, the total number of pages in the directory was divided by the number of students in that element. Every Qth page was then designated as a page for selecting one student for the given school and the given class. The rule was: select the first student whose name appears on that page who satisfies the class and school requirement.

In selecting like alumni for participation in the study, the Bulletin of Purdue University for 1936–37 was used, since it listed the names of students by class and school enrolled for that year. Using the same system of selecting alumni as was used for selecting students, the problem of obtaining the proper number of alumni drop-outs at various stages of their academic careers to make the student and alumni samples comparable did not have to be met. We simply assumed the students in attendance in 1936–37 were similar in ability to students in attendance in 1948–49. The Alumni Association provided addresses for the alumni (graduates and non-graduates) we had listed.

The Rating Scale and Instructions.—The Rating Scales and instructions were mailed to students and alumni alike under the accompanying covering letter.

PRESIDENT'S OFFICE
Purdue University
Lafayette, Indiana

February 28, 1949

Dear Alumnus or Student:

Purdue's standing among educational institutions depends largely on the quality of the teaching that goes on in its classrooms and lecture halls. There are, however, many unexplored areas of the problem "What makes a good teacher?" Attempts to supply an answer have included asking present students to judge their teachers, but such attempts have frequently been attacked on the grounds that the undergraduate student lacks the maturity for supplying valid judgments of this sort. It is proposed to investigate this type of criticism more closely with the help of students and alumni.

As students, will your answer to the question "What makes a good teacher?" change as you grow older? Or, as alumni, have your ideas as to what makes a good teacher changed since your student days? These are two aspects of the question we wish answered. To help us find the answer I am calling upon you to coöperate in an investigation being made by the Division of Educational Reference. Your task, as explained in the enclosures, is quire simple and brief. It includes rating your past or present teachers at Purdue University. I wish to urge you to complete it thoughtfully and at your earliest convenience.

I should like to assure you that *results of the study will never be identified with any specific instructor by any member of the instructional staff.* The ratings will be seen only by one or two confidential clerks. Your ratings will therefore not affect any instructor's standing. All the teachers listed on the attached form have individually agreed to participate in the investigation.

It will be obvious to you, I am sure, that your evaluation of instructors should be entirely your own. It is most important, therefore, that you make these evaluations without discussing them with anyone.

I shall be grateful for your coöperation in this study.

Sincerely yours,
Frederick L. Hovde
President

The student or alumnus was first asked to examine the list of traits appearing on an enclosed copy of the Purdue Rating Scale for Instructors and rank the traits from 1 to 10 as to their importance to good college teaching. A brief questionnaire was then provided requesting sex, age, year of graduation, total number of semesters in attendance at the University, school or division, current semester (if a student) and veteran or non-veteran status. Next the student or alumnus was asked to scan the list of one hundred two instructors' names and to check the names of those from whom he could recall having had courses for at least five weeks of a semester. He was instructed how to select at random no more than five of those instructors for rating. Rating of all five instructors was done in the traditional manner[3] for all ten traits, except that all five ratings were placed on the same scale for each trait with the instructor identified by a number. The rating scale and questionnaire were then returned by mail.

The Returns.—Of one thousand questionnaires and Rating Scales sent to students and alumni, two hundred fifty-one students and one hundred thirty-eight alumni submitted ratings in time for the analysis. The students returned a total of eight hundred thirty ratings and the alumni a total of five hundred seventy-seven ratings. The composition of the student and alumni groups which sent in ratings can be seen in Table 2.

From Table 2 we can surmise that the samples returning ratings of instructors are fairly similar with regard to sex and school or department. Moreover, the proportions of response by both students and alumni by schools and departments were approximately similar to the proportions in the original mailing. The larger number of veterans in the alumni group is to be expected, as are, of course, the differences in age. The most striking difference between students and alumni returning ratings, however, is found in the number of semesters completed. It will be recalled that alumni were selected from a listing (student directory of thirteen years ago) including names of those who would fall by the wayside, scholastically, or would leave school for other reasons. Yet, apparently, almost none of the drop-out alumni receiving questionnaires returned them. We must admit, therefore, the possibility that the alumni group may be slightly superior academically, notwithstanding the fact that the student group is heavily loaded with juniors and seniors.

The number of ratings received for any individual instructor by students or by alumni does not exceed fifty. In a few cases no students rated a given instructor and in a few more cases no alumni gave a given instructor ratings. Analysis of results to make comparisons between ratings of instructors and of traits by alumni and students have to be confined to ratings made as a whole on all instructors as a group.

TABLE 2.—CHARACTERISTICS OF STUDENTS AND ALUMNI RETURNING RATINGS

	Students	Alumni
Males	80%	84%
Females	20%	16%
Mean Age	21.8 yrs.	33.0 yrs.
Age Range	18–31	30–50
School, Division or Department		
Trade & Industrial Education	1%	2%
Engineering	22%	28%
Science	43%	34%
Pharmacy	5%	5%
Home Economics	4%	5%
Agriculture	22%	24%
Physical Education	3%	2%
Semester of Work Completed		
1	0%	
2	9%	
3	3%	
4	18%	
5	9%	1%
6	22%	
7	11%	1%
8	28%	2%
Graduated	0%	97%
Veteran, with Overseas Service Before September, 1945	26%	38%

RESULTS

Desiderata of Good Instructors.—To compare the relative importance to alumni and to students of the ten traits on the

Purdue Rating Scale for Instructors, the median rankings of each trait by the two hundred fifty-one students and the one hundred thirty-eight alumni were obtained. The Pearson product-moment correlation coefficient between these medians for the ten traits was .92. On the basis of the obtained medians the rank order correlation was also computed to be .92, either method indicating a strong degree of relationship between relative importance attached to the various rating traits by students and alumni. In Table 3 one can see that seven of ten traits have identical rankings by students and alumni. Trait 2, 'Sympathetic Attitude Toward Students,' is ranked fifth in importance by students and sixth by alumni. Trait 3, 'Fairness in Grading,' is ranked sixth by students and eighth by alumni. Trait 6, 'Sense of Proportion and Humor,' shows the greatest discrepancy between students' ranking and alumni ranking in that it is ranked eighth by students but fifth by alumni. Table 3 lists the traits in the order of importance by student ranking and gives the median rankings by students and alumni.

The difference in emphasis placed on such traits as 'Fairness in Grading' and 'Sense of Proportion and Humor' by alumni and students lends itself readily to speculation. Grades unquestionably assume their greatest importance to the individual when set as immediate goals but the alumnus, as his academic days recede, must gradually realize the over-emphasis placed upon grades in their relation to later success. As for 'Sense of Proportion and Humor,' the student is possibly too much concerned with his academic progress, has too much seriousness of intent to place much premium upon this trait. The alumnus, on the other hand, possibly has realized the importance of fun and humor to individual adjustment and has learned how to relax when relaxation is needed.

These trait rankings compare favorably, although not perfectly, with a much earlier study in which another group of students ranked the traits.[4] In general these students showed a high degree of agreement in their judgments of the relative importance of the ten traits, with the exception of 'Fairness in Grading.' The highest three in order of relative importance were 'Presentation of Subject Matter,' 'Stimulating Intellectual Curiosity,' and 'Interest in Subject.' The lowest were 'Personal Peculiarities,' 'Personal Appearance' and 'Sense of Proportion and Humor.'

Comparison of Student-Alumni Ratings of Instructors.—In obtaining a comparison between student and alumni ratings of the same instructors two types of analysis appeared desirable—that which would give a measure of the degree of relationship between the ratings and that which would give a measure of the amount and direction of any average difference that occurred for a given trait. We will first discuss the correlations obtained between student and alumni ratings of the same instructors for each trait.

TABLE 3.—RANKING OF PRSI TRAITS BY STUDENTS AND ALUMNI

Trait	Student Rank Order	Alumni Rank Order	St Mdn	Al Mdn
Presentation of Subject Matter	1	1	.9	1.3
Interest in Subject	2	2	2.1	2.0
Stimulating Intellectual Curiosity	3	3	3.5	2.5
Liberal and Progressive Attitude	4	4	4.1	4.7
Sympathetic Attitude toward Student	5	6	4.6	5.3
Fairness in Grading	6	8	4.8	6.1
Self-Reliance and Confidence	7	7	5.6	5.8
Sense of Proportion and Humor	8	5	6.0	5.0
Personal Appearance	9	9	8.7	8.7
Personal Peculiarities	10	10	8.7	8.8

In obtaining correlations between student and alumni ratings for instructors it was necessary to correlate the sum of a given number of alumni ratings with the sum of the same number of student ratings for each instructor. This procedure was repeated for ten traits, giving in all ten correlation coefficients. Here the paucity of our data first became apparent. We had the choice of basing our correlations upon the sum of two student ratings against two alumni ratings for the same instructor (in which case we would have seventy-seven out of the one hundred two instructors represented in the comparison) or of basing the correlations upon the sum of eight student ratings against eight alumni ratings which would enable us to use only seventeen out of the one hundred two instructors. Although the number of cases of instructors entering the correlation would be only seventeen, it was decided to use the latter method because of the superior reliability of the variables (the sum of eight ratings instead of only two). In an early study it was found that ratings by from ten to twenty students on a single trait for instructors "differing sensibly in the amount of the trait possessed yield reliabilities which compare favorably with the reliabilities reported for standardized mental and educational tests."[8] In a recent investigation ratings based on the average of ten ratings, randomly selected, of two hundred five instructors, correlated against ten other ratings of the same instructors, randomly selected, had reliabilities ranging from .72 to .87.[1]

TABLE 4.—ESTIMATION OF RELIABILITY OF PRSI TRAITS BASED ON THE CORRELATION OF EIGHT RATERS VERSUS EIGHT OTHER RATERS

Trait No.	Reliability[1] Based on 10 × 10 Ratings	Reliability Estimated for 8 × 8 Ratings
1	.86	.83
2	.76	.72
3	.72	.67
4	.75	.71
5	.87	.84
6	.78	.74
7	.83	.80
8	.82	.79
9	.83	.80
10	.84	.81

For this study, reliabilities of the average of eight randomly selected ratings against eight other randomly selected ratings were estimated from the above study[1] by the Spearman-Brown Formula. These estimated reliability coefficients, shown in Table 4, vary from .67 to .84. This is evidence that the measures of student and alumni judgments that comprise the variables in our relationships are fairly reliable. Of course, to the extent that these measures are not reliable, the obtained relationships shown in Table 5 between student and alumni ratings are lower than they would be with a larger number of ratings per instructor.

As can be seen in Table 5, the correlation coefficients range from .40 to .68, indicating a substantial correspondence between student and alumni ratings of the same instructors. These measures indicate that to a large extent the same relative ratings are given to instructors by current students whose average age is about twenty-two years and by alumni with an average age of thirty-three years.

TABLE 5.—COEFFICIENTS OF CORRELATION BETWEEN THE SUM OF EIGHT STUDENT RATINGS AND THE SUM OF EIGHT ALUMNI RATINGS OF SEVENTEEN INSTRUCTORS FOR TEN PRSI TRAITS

Traits	*r*
1. Interest in Subject	.40
2. Sympathetic Attitude toward Students	.68
3. Fairness in Grading	.47
4. Liberal and Progressive Attitude	.59
5. Presentation of Subject Matter	.58
6. Sense of Proportion and Humor	.46
7. Self-reliance and Confidence	.64
8. Personal Peculiarities	.50
9. Personal Appearance	.65
10. Stimulating Intellectual Curiosity	.46

Next it was desirable to investigate the actual differences that occurred in comparing ratings by students and alumni. For this analysis it was possible to use most of the ratings received, or all the ratings on ninety-two of the one hundred two instructors on whom we had at least one rating by both students and alumni. The procedure for each of the ten traits was simply to compute, regardless of the number of raters, the mean alumni and student rating received by each instructor as representative of student or alumni judgment concerning that instructor. These mean ratings were then averaged for the ninety-two instructors and the average differences tested by computing the standard error of the differences between means, taking into account the correlation between the measures. Based on the data from this investigation, a significant difference between these averages for a given trait would then represent a true tendency for students or alumni to rate instructors higher or lower by a given amount.

On three traits, 'Interest in Subject,' 'Sense of Proportion and Humor' and 'Self-Reliance and Confidence,' students tend to rate instructors higher than do alumni. These differences of 4 to 5 points are significant at the ninety-nine per cent level of confidence. On one trait, 'Fairness in Grading,' alumni tended to rate instructors about 3 points higher than did students. This difference was significant at the ninety-five per cent level of confidence. Complete results of this type of analysis are given in Table 6. In addition to the above significant differences which we have pointed out, there appears a slight tendency for students to give slightly higher ratings than alumni on most traits. Alumni are also considerably more variable in their ratings on Traits 1, 7 and 9.

The tendency for students to rate instructors lower than do alumni on 'Fairness in Grading' may possibly reflect again the abnormal concern of undergraduates with grades and grading methods. For the rest of the traits, however, either there is no significant difference or the students tend to rate instructors slightly higher than do alumni.

Teacher criticisms of the Purdue Rating Scale for Instructors made on grounds of student immaturity probably are founded upon the belief that undergraduates are not likely to give instructors their due. Our results suggest otherwise; students of this study have tended to rate the instructors of this study higher than have alumni on two traits but not much differently than alumni on the other eight traits. On the other hand, superior ratings by students may reflect just those changes that have occurred in teaching during the past decade, leaving the distinct possibility that there actually are no differences in the way alumni and undergraduates rate the same instructors. In all fairness, however, we should consider the possibility that the change for the better in teaching ability of the ninety-two instructors was so great that these slightly superior student ratings actually are under-ratings of a superior teaching ability.

SUMMARY AND CONCLUSIONS

To test the hypothesis that ratings of instructors do not change with the maturity of the rater, comparisons were made of the ratings of ninety-two Purdue University instructors by one hundred thirty-eight alumni of at least ten years' standing with ratings

TABLE 6.—DIFFERENCES IN AVERAGE MEAN RATINGS BY STUDENTS AND ALUMNI OF 92 INSTRUCTORS

Trait	Average Mean Rating Students	Average Mean Rating Alumni	Sigmas Mean Rating Students	Sigmas Mean Rating Alumni	Differences in Average Mean Ratings	SE of Diffs	Critical Ratios
Interest in Subject	78.6	74.8	8.6	11.0	3.8	1.32	2.88*
Sympathetic Attitude	68.5	67.0	14.2	14.2	1.5	1.64	.91
Fairness in Grading	71.8	75.1	10.8	9.8	−3.3	1.42	2.32†
Liberal and Progressive Attitude	65.4	64.3	13.3	12.9	1.1	1.58	.70
Presentation of Subject	63.8	63.5	14.5	14.2	.3	1.83	.16
Sense of Proportion and Humor	64.5	59.8	13.4	13.6	4.7	1.54	3.05*
Self-reliance and Confidence	73.7	68.6	11.0	13.6	5.1	1.67	3.05*
Personal Peculiarities	68.5	66.5	11.7	12.9	2.0	1.70	1.18
Personal Appearance	78.4	75.6	9.2	11.6	2.8	1.55	1.81
Stimulating Intellectual Curiosity	60.7	61.3	12.6	12.2	− .6	1.55	.39

* Difference is significant at the ninety-nine per cent level of confidence.

† Difference is significant at the ninety-five per cent level of confidence.

by two hundred fifty-one current undergraduates. The relative importance of the ten traits of the Purdue Rating Scale for Instructors to alumni and students was also examined. The following tentative conclusions appear justified concerning the comparative ratings of Purdue University instructors by current students and alumni of ten to thirteen years ago:

1) There are substantial positive relationships between the relative average ratings of instructors by students and alumni.

2) The average differences that do occur between student and alumni ratings of Purdue University instructors indicate that present students as a group rate their instructors slightly higher than they will as older individuals, but that such differences are actually too small to be of practical significance.

3) Purdue University students and alumni agree highly upon the relative importance to good instruction of the ten traits of the Purdue Rating Scale for Instructors.

These conclusions hinge upon the satisfying of the general assumption that the selections of students, alumni and instructors did not account for the results obtained. In particular, results in favor of student ratings over alumni ratings for the same instructors may reflect improvements in teaching traits of the ninety-two instructors over the period covered by the study.

This study, then, has presented evidence that the judgments made by undergraduates of their instructors are valid ones in terms of permanence and maturity. Instructors should be able to avail themselves of the Purdue Rating Scale for Instructors with increased confidence that the judgments they receive thereby are worth-while ones in evaluating their own teaching personalities.

BIBLIOGRAPHY

1) Elliott, D. N. *Characteristics and Relationships of Various Criteria of Teaching*. Unpublished Ph.D. thesis, Purdue University, 1949.

2) Goodhartz, A. S. "Student Attitudes and Opinions Relating to Teaching at Brooklyn College." *School and Society*, 1948, 68, 345–349.

3) Remmers, H. H., and Brandenburg, G. C., "Experimental Data on the Purdue Rating Scale for Instructors," *Educational Administration and Supervision*, Vol. XIII, No. 8, pp. 519–527, November, 1927.

4) ———, and Stalnaker, J. M., "Can Students Discriminate Traits Associated with Success in Teaching?" *Journal of Applied Psychology*, Vol. 2, No. 6, pp. 602–610, December, 1928.

5) ———, "The College Professor as the Student Sees Him," *Bulletin of Purdue University, Studies in Higher Education XI*, Vol. xxix, No. 6, p. 75, March, 1929.

6) ———, and Wykoff, G. S., "Student Rating of College Teaching—A Reply," *School and Society*, Vol. xxx, No. 764, pp. 232–234, August, 1929.

7) ———, and Elliott, D. N. "The Indiana College and University Staff Evaluation Program." *School and Society*, 1949, 70, 168–171.

8) ———, "The Equivalence of Judgments and Test Items in the Sense of the Spearman-Brown Formula," *J. educ. Psychol.*, 1934, 18, 619–630.

SOCIAL PERCEPTION AND TEACHER-PUPIL RELATIONSHIPS

N. L. GAGE and GEORGE SUCI

Bureau of Research and Service
College of Education
University of Illinois

How two or more persons interact in a given situation depends on many variables, only a few of which have been isolated even conceptually. Properties of the persons and of their situations, all considered as interdependent aspects of a field of forces, obviously provide handles by which to grasp the interaction phenomena. This paper seeks to throw light on interactions between teachers and pupils as a function of one property of the teachers: the accuracy of their perception of pupils' attitudes.

We are concerned with the more affective dimensions of the interaction, rather than with cognitive ones. Our reasoning goes as follows: Behavior in any situation is a function of one's perception of the situation. The 'appropriateness' of the behavior will hence depend, at least in part, on the correctness or accuracy of the perceptions. But mere perceptions do not govern behavior completely; the motivational system of the person will determine the ends to which his perceptions are used. Similarly, his 'integration,' or freedom from neuroticism, will determine how well he can control his behavior to accord with his motivations and perceptions. Thus, even if John perceives a stop light accurately, he will not stop for it unless he wants to and is sufficiently unexcited to be able to react in time. If John perceives his date's boredom with a movie, he will not suggest that they leave unless he both wants to please her and this motive is integrated in a given way with his own feelings about the movie.

So far, however, this formulation ignores the relationships between these determiners; motivation and perception, at least, are themselves linked together. As Bruner[1] recently summarized it:

> "(a) Personally relevant objects in the perceptual field, whether the relevance is a function of positive or negative value to the individual, undergo accentuation. . . .
>
> "(b) Perceptual selectivity . . . is shown to favor the

recognition of valued objects and objects associated with the prevailing interests of the subject."

Applied to teachers' perceptions of pupils' attitudes and their interaction with pupils, this implies that the accuracy of teachers' perceptions will be a function of the way in which teachers value pupils and relationships with them. Miss Smith will tend to become aware of how her pupils feel about various issues to the degree that she is interested in those pupils and in getting along smoothly with them. Similarly, she will probably act upon her awareness of their attitudes in positive ways for the same reasons that motivated her to perceive them accurately. Thus certain patterns of motivation, perception and behavior tend to go together. At the one extreme of a continuum, we may have teachers who do not value pupils positively, do not perceive their attitudes accurately, and behave in ways conducive to negative reactions. At the other extreme are teachers who 'need' their pupils more, perceive them with greater sensitivity and understanding, and conduct themselves in ways that elicit positive affect from pupils.

But, it may be objected, need does not necessarily lead to accuracy of perception. The more an object is valued, the more indeed may perception of it be distorted. In our framework, this leads us to examine what it is we are asking teachers to perceive. This, as will be described below, was the attitudes of pupils as reported by themselves in answer to direct Yes-No questions. Whether the attitudes so reported are 'true' or 'real' ones, we cannot know. It is reasonable to suppose, however, that what pupils report about themselves in a nonclinical, classroom situation are probably opinions of an ego-supporting kind. The opinions of the pupils tended, in other words, to be those which would enhance their own view of themselves. When teachers are then asked to estimate these opinions, they are being asked to judge what are the ego-supporting responses of pupils. Teachers who value pupils positively then should come closer to the pupils' own picture of themselves. Such a 'coming closer' is what we have here termed 'accuracy of social perception.'

In short, this should mean that teachers who perceive their pupils' attitudes more accurately should be regarded more favorably by their pupils. To test this inference, we need estimates of (1) the accuracy of teachers' perceptions of pupils' attitudes,

(2) the favorableness of pupils' attitudes toward their teachers, and (3) the relationship between (1) and (2).

THE ESTIMATE OF ACCURACY OF SOCIAL PERCEPTION

Ideally, we should deal with the teacher's perception of attitudes of individual pupils on matters that are deeply relevant to each pupil. This would mean, first, the identification of attitude-objects that are highly salient for each pupil and, second, asking the teacher to estimate that pupil's attitude toward those objects. The attitudes to be perceived would then differ from one pupil to the next. Such a procedure was manifestly too difficult for the present study. We compromised by selecting areas of attitude that we surmised to be significant to the pupils *en masse*, and we asked the teachers to estimate the group attitude on each topic. This procedure is admittedly gross; but whatever support for our hypothesis might emerge from such an approximation might then be considered an underestimate of its true validity.

Accordingly, we proceeded as follows: The twenty teachers of a high school were asked to estimate what per cent of the two hundred pupils in the school would respond affirmatively to a set of sixty-seven opinion items. The items dealt with scholastic, recreational, and student government issues in their school.* An illustrative item of each kind follows:

Scholastic: "Should students with high scholastic ratings be allowed to skip classes?"

Recreation: "Could the time spent in school activities be used better for studying or working?"

Student government: "Should Student Council members be appointed by teachers rather than elected by students?"

The students responded anonymously, underlining 'Yes' or 'No' for each item. The pupils' 'Yes' responses were tabulated and the percentage of Yes's found for each item. These percentages comprised the scoring key. Each teacher was scored by taking the difference between his estimate and the actual percentage for each item, summing, and, because a few teachers omitted some items, averaging over-all items. The signs of the differences were disregarded. Such 'mean error' scores were also obtained separately for each of the three subgroups of items.

---

* These items were adapted from those used by Wood.[4]

These 'mean error' scores are directly dependent on the absolute accuracy with which each percentage is predicted. But a teacher, although not able to predict accurately in terms of absolute percentages, may be able to rank the items in order of relative acquiescence of the pupils. Accordingly, we also computed for each teacher a Pearsonian *r* between actual and predicted percentages. This we called an '*r*-score.' It correlated .77 with the 'mean error' score, showing a possibly important difference between the two scoring methods.

PUPILS' INTERACTION WITH THEIR TEACHERS

At a separate meeting, with the teacher absent, the pupils rated their six current teachers on fifty-two Yes-No-? items* typified by the following:

"8. Is this teacher often 'bossy?'"

"14. Does this teacher think he or she is always right and the student wrong?"

Again the papers were unsigned and pupils were assured their responses would remain anonymous. A score for each teacher was obtained by scoring favorable and unfavorable pupil responses as 1 and 0, respectively, summing over-all items and averaging over-all pupils rating the given teacher. This measure we considered an approximation, adequate for this study, to the positive affect elicited in pupils by their interaction with a given teacher.

Furthermore, for an appraisal of each teacher's system of values in teacher-pupil relationships, we used the *Teacher Attitude Inventory* developed by Leeds and Cook.[3]† The two hundred thirty-nine items in this device had been empirically selected and weighted by Leeds and Cook for their ability to discriminate between teachers known to maintain very satisfactory and very unsatisfactory relations with pupils, in the judgment of the teachers' local school administrators.

RESULTS

Our first task in analyzing the data was to estimate the generality over items, or the internal consistency, of our approach to accuracy to social perception. Table 1 shows the corrected odd-even coefficients for the 'mean error' scores on all items and on each of the three sub-groups. The coefficient for all items, .73, indicates that the accuracy of teachers in estimating percentages of acquiescence was fairly general over the items. Thus we have evidence that the task of judging opinion revealed individual differences among teachers that were substantially loaded with non-error variance. For the subgroups of items, these estimates yielded much lower coefficients.

---

* These were a slight modification of those used by Leeds and Cook.[3]

† We are grateful to Professor W. W. Cook for permission to use this device.

TABLE 1.—CORRECTED ODD-EVEN RELIABILITY ESTIMATES FOR THE 'MEAN ERROR' SCORES ON OPINION ESTIMATES

$N = 20$

Type of Items	No. of Items	Odd Items		Even Items		Corrected Odd-Even r
		Mean	SD	Mean	SD	
All Items 1–67	67	20.19	2.81	18.64	2.97	.73
Scholastic 1–23	23	22.50	5.03	19.74	4.05	.17
Recreation 24–45	22	15.47	3.13	18.16	4.28	.28
Student Government 46–67	22	20.50	5.07	17.26	4.95	.47

Secondly, what is the reliability of the pupils' ratings of their teachers? Table 2 shows the results obtained from an application of Horst's formula.[2] This coefficient, .93, indicates that the mean ratings we obtained would correlate very highly with subsequent sets of similar ratings, when our estimate of the variability of such further mean ratings is based on an average of the standard errors of those obtained. In short, the variance among the mean ratings is even more a non-error one; the ratings discriminated well among the teachers.

Having established these estimates of reliability, we turned to the relationships with which we were primarily concerned. Is social perception, defined operationally as mean error in estimating students' percentages of acquiescence to opinion items, related to acceptance of the teachers by pupils, defined as the mean rating of teachers by pupils? Table 3 shows the pertinent coefficients of correlation. The *r's* between mean rating and mean error for all items and for the three subgroups of items are $-.37$, $-.16$, $-.30$, and $-.18$, respectively. It is evident that all *r*'s are in the direction inferred from our hypothesis: the greater the teacher's mean error in estimating student opinion, the lower his mean rating by students. None of these *r*'s are statistically significant. Our interpretation of this fact may, however, be tempered somewhat by the consistency in direction and size of the three subtest coefficients. These subtests provide in a sense three semi-independent trials of our investigation. The consistency of their results may therefore be construed as favorable evidence in addition to that provided by the over-all test's *r*.

TABLE 2.—ESTIMATE BY HORST'S METHOD OF THE RELIABILITY OF PUPILS' RATINGS OF THEIR TEACHERS

$$r = 1 - \frac{\sum \frac{\sigma_i^2}{n_i - 1}}{N\sigma_{M_i}^2} = 1 - \frac{47.45}{648.46}$$

$$= .93$$

where N = number of teachers
n_i = number of pupils rating teacher i
M_i = mean of the ratings for teacher i
σ_i = standard deviation of ratings for teacher i
σ_{M_i} = standard deviation of means for the N teachers

When the teachers' estimates of student opinion were scored in terms of their correlation with, rather than their difference from, the actual percentages, we obtain the results also shown in Table 3. Here the relationships, using r and z_r as measures of accuracy of social perception, increase to .50 and .45, respectively. Both these coefficients meet the requirement for statistical significance at the five per cent level.

Why do the *r*-scores correlate more highly than the mean error scores with the pupils' ratings? We can only conjecture as follows: The *r*-scores are not influenced, as are the mean-error scores, by individual differences among teachers in 'adaptation

TABLE 3.—MEANS, STANDARD DEVIATIONS AND INTERCORRELATIONS* OF VARIOUS MEASURES

	No. of Teachers	Mean	SD	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1. Mean Rating by Pupils	20	33.20	5.49	−.37	−.16	−.30	−.18	.50	.45	−.20
2. Mean Error (All Items)	20	19.35	2.45					−.77		−.57
3. Mean Error (Scholastic Items)	20	21.65	3.54			.13	.18			
4. Mean Error (Recreational Items)	20	16.70	3.28				.48			
5. Mean Error (Student Government Items)	20	19.15	4.60							
6. r-score (All Items)	20	.67	.08							.46
7. z_r-score (All Items)	20	.82	.16							
8. Teacher Attitude Inventory	19	154.08	18.35							

* When $N = 20$, an r of .44 is significant at the five per cent level.

level,' or 'anchor-point.' The teachers differed very significantly, beyond the one per cent level, in the means of their predicted percentages. The *r*-scores eliminate the effect of these differences. They render the measure of accuracy in social perception dependent solely on relative rather than absolute judgments of degree of acquiescence.

Two further findings should be mentioned. First, as shown in Table 3, the scores on the Cook-Leeds inventory correlated .57 and .46 with our mean error and *r*-scores, respectively. This indicates that the attitudes and understandings concerning pupils which are tapped by that test are significantly related to ability to estimate students' opinions. But what of the relation of the Cook-Leeds instrument to mean ratings by pupils? Contrary to the Cook-Leeds findings, our coefficient was negative, −.20, although not significantly different from zero. In this school, at any rate, this test did not yield results in corroboration of the significant positive relationships with pupils' ratings (about .45) previously reported by Cook and Leeds.

DISCUSSION

All our conclusions must be tempered by an awareness of the smallness of our sample and its lack of replication in other schools. For example, our impression that a curvilinear relationship exists between pupil ratings and accuracy of perception cannot adequately be pursued. It should be reported, however, that six teachers with high mean ratings nevertheless made only mediocre 'mean error' or '*r*' scores. None of the teachers with very low ratings made high accuracy in social perception scores. This could have occurred if some of the well-liked teachers did not attack the task of estimating pupils' opinions with sufficient motivation. Thus low ratings seem sufficient to indicate poor perception scores, but the converse is not true.

The instability of our results is further indicated when we omit one extreme case from the calculations of the *r*'s between the two scoring methods and the mean ratings. With the one case retained, *r*'s of −.37 and .50 were obtained with the 'mean error' score and the '*r*-score,' respectively. With the extreme case omitted, the *r*'s became −.24 and .18, respectively—a considerable reduction. Nevertheless some degree of correlation, in the expected direction, persisted.

SUMMARY

To test the hypothesis that accuracy of social perception is positively related to effectiveness of interpersonal relations, we asked the twenty teachers of a high school to predict the percentage of the two hundred students who would answer 'Yes' to each of sixty-seven items eliciting opinions on various aspects of the school. Teachers' predictions were scored in terms of (a) 'mean error' from actual percentages and (b) correlation with actual percentages. The teachers were rated by their pupils on a fifty-two-item inventory. Finally, the teachers filled out the Cook-Leeds *Teacher Attitude Inventory*.

Results support the hypothesis. The teachers' 'mean error' scores had fair internal consistency (corrected odd-even r, .73) and correlated $-.37$ with mean rating by pupils. Their 'r-scores' correlated .50 with mean rating. The teachers' scores on the *Teacher Attitude Inventory* correlated .57 with the 'mean error' scores but, unexpectedly, $-.20$ with mean ratings. (For $N = 20$, an r of .44 is significant at the five per cent level.)

Our results must, of course, be pursued through replications in other schools and other situations. Tentatively, however, we conclude that teachers' accuracy of social perception is positively related to their effectiveness in eliciting positive affect in pupils.

REFERENCES

1) J. S. Bruner. "Social psychology and group processes." In C. P. Stone and D. W. Taylor, (Eds.) Annual Review of Psychology, Vol. I, 1950, pp. 119–150. Stanford, California: Annual Reviews, Inc.

2) Paul Horst. "A generalized expression for the reliability of measures." *Psychometrika*, Vol. XIV, 1949, pp. 21–31.

3) C. H. Leeds and W. W. Cook. "The construction and differential value of a scale for determining teacher-pupil attitudes." *Journal of Experimental Education.*, Vol. XVI, 1947, pp. 149–159.

4) Homer Wood. *An Analysis of Social Sensitivity*. Unpublished Ph. D. Thesis, Yale University, 1948.

THE READING EASE, HUMAN INTEREST VALUE, AND THEMATIC CONTENT OF *ST. NICHOLAS MAGAZINE:* A STUDY OF CHILDREN'S LITERATURE

BARBARA SWARD and DALE B. HARRIS
Institute of Child Welfare, University of Minnesota

Perhaps the best known and most widely used technique for the analysis of written prose has been proposed by Flesch.[2,3,4] His formulae are extensively used by journalists, writers of articles for trade magazines, and others. Although many techniques for the analysis of educational material prepared for children have been developed, little work has been done on prose specifically designed for children's entertainment.

Anyone professionally concerned with child development or education is familiar with the impression that stories and books for children have improved greatly in recent years. There is a widespread impression that children's literature of the Victorian era was stilted, artificial, highly moralistic. The writers, interested in this subject and in the adaptation to children which has presumably occurred in the informed adult world as a result of scientific study of children, here attempt an analysis of two problems. They are interested in the 'readability' of selected children's material of an earlier day and of the present. They are also interested in the ideological content of this material, and how it may have changed over the years, due to changing conceptions both of children and of social values.

For the first problem, the Flesch formulae suggest themselves as convenient indices of 'Reading Ease' and of 'Human Interest.' The formulae, their rationale and derivation are available and will not be discussed here. For the comparative purposes of this paper, two monthly children's magazines were selected. *St. Nicholas*, the 'grandfather' in its field, commenced publication in 1873 under the editorship of Mary Mapes Dodge. It enjoyed wide popularity until it ceased publication in 1940. In an earlier day, its pages were crowded with the work of such noted writers as Louisa May Alcott, Kipling, Twain, Bret Harte, Thomas Bailey Aldrich, William Cullen Bryant and others. *Child Life*, which, of the modern children's magazines, has the longest continuous publication and the widest circulation, commenced publication in 1922 and enjoys considerable repute today.

To obtain materials for analysis, the writers proceeded as follows: Samples of prose from both magazines were taken from issues published in successive decades. In order to utilize the first issue of *St. Nicholas*, the November, February, May, and August issues of the year 1873–4 were selected to yield the first sample. Comparable issues of succeeding decades gave additional samples. The same months were used for the *Child Life* series, but, because a complete file was unavailable to the writers, the analysis began with the 1925–26 period. Thus, counting both magazines, a seventy-three year period is covered, 1873 to 1946, inclusive.

For each selected monthly issue of each magazine (twenty-eight issues for *St. Nicholas* and twelve for *Child Life*), the first, third, and fifth fictional stories provided the material for analysis. From each story, two one-hundred word samples were taken, one sample beginning with the third paragraph and one sample beginning with the fifth paragraph.* The first, third, and fifth stories were used consistently except where this procedure would include a succeeding installment of a serial story already sampled, or another story by an author already sampled; then, the story immediately following the designated story was used.† Few, if any, of the monthly issues contained more than five or six such fictional stories.

In each case, a fictional story was considered to be a story that was pure fiction or fictionalized history in prose (many of the Washington and Lincoln stories for children are of this latter type), and of at least a page or more in length. Plays were not used because extensive dialogue would make the 'personal sentences' percentages in the Flesch formula spuriously high. Serial stories were used, but in counting paragraphs, the synopsis section was ignored as being abbreviated and very non-personal.

Flesch's Reading Ease (R.E.) score is easy to compute and quite objective. It is based on the average number of words per sentence, and the average number of syllables per one hundred words. The Human Interest (H.I.) score involves greater subjectivity of judgment. It involves the number of 'personal words' and the number of 'personal sentences' per sample of one hundred words. 'Personal words' are not always self-evident. The writers encountered many nouns which could not be said objectively to possess masculine or feminine gender, yet because of their functional use in the story it seemed wise to consider them such. Examples are: Lady Hollyhock, Buster Bumblebee, Kipling's mongoose Rikki-Tikki-Tavi, etc. The rule adopted was that if non-human characters were definitely personalized and possessed gender, nouns designating them were counted as personal words. If they were neutrally or realistically portrayed, they were not so categorized.*

---

* We selected the third and fifth paragraphs in an effort to cover adequately two types of writing. The third paragraph frequently caught the introductory variation or exposition, while the fifth usually included more dialogue or light, personalized writing.

† We avoided using more than one story by the same author because we were interested in the variety of writers representing a particular period.

TABLE I.—MEDIAN AND SEMI-INTERQUARTILE RANGE VALUES FOR READING EASE SCORE DISTRIBUTIONS FOR SELECTED YEARS

Magazine and Year	Median R.E. Score	Q	Flesch Description of Style
St. Nicholas			
1873–4	79.4	5.3	Fairly easy
1883–4	76.7	7.7	Fairly easy
1893–4	80.7	9.4	Easy
1903–4	80.9	6.2	Easy
1913–4	80.5	5.6	Easy
1923–4	75.5	6.0	Fairly easy
1933–4	80.4	8.3	Easy
Child Life			
1925–6	75.6	7.3	Fairly easy
1935–6	84.6	4.7	Easy
1945–6	78.4	6.3	Fairly easy

The median R.E. scores for *St. Nicholas* and *Child Life* are presented in Table I. The descriptions of style in the table are borrowed from Flesch's designations for adult material.

---

* This rule, that in deciding 'personal words' the functional use of the word in the paragraph, not necessarily the word's absolute or commonest use, was applied throughout. Flesch does not state this rule. The problem does not occur in the newspaper studies where the Flesch technique has been most widely used. It is a definite problem in imaginative material prepared for children.

The outstanding result of this analysis is the fairly uniform level of reading ease of the material over three-quarters of a century. One might expect that the stress in recent decades on simplified, pictorialized literature for children might have brought with it a somewhat higher typical R.E. score derived from the prose. The great change in format of children's material is apparent to all, but the early issues of *St. Nicholas*, with their few, black-and-white pictures, and small print, were just as 'easy' in prose as the modern children's magazines.

TABLE II.—MEDIAN AND SEMI-INTERQUARTILE RANGE VALUES FOR HUMAN INTEREST SCORE DISTRIBUTIONS FOR SELECTED YEARS

Magazine and Year	Median R.E. Score	Q	Flesch Description of Style
St. Nicholas			
1873–4	51.6	19.7	Highly interesting
1883–4	71.9	26.8	Dramatic
1893–4	72.4	27.9	Dramatic
1903–4	55.9	19.7	Highly interesting
1913–4	56.5	18.0	Highly interesting
1923–4	64.2	18.7	Dramatic
1933–4	64.7	15.3	Dramatic
Child Life			
1925–6	57.6	19.9	Highly interesting
1935–6	64.4	15.0	Dramatic
1945–6	57.0	21.2	Highly interesting

Human interest scores are presented in Table II. The apparently greater variability among these medians must be interpreted in the light of the greater basic variability in the score itself. The greatest difference—that between the years of 1873–4 and 1893–4—is significant only at the five per cent level, statistically; that is, the chances are still five in one hundred that a difference as large or larger than the observed one could occur from 'errors' of sampling. Again, the high interest value of children's literature in another era must be noted. Small wonder that Henry Steele Commager's anthology* from *St. Nicholas* has been so well received.

---

* Henry Steele Commager (Editor), *The St. Nicholas Anthology*. New York, Random House, 1948. According to the publishers, approximately fifty thousand copies of this anthology have been sold.

Two additional analyses are perhaps of sufficient interest to note the general results in passing without presenting tabular detail. The reading ease of children's material results more from the selection of short, simple sentences. An analysis of mean sentence length reveals the interesting fact that, judged by sentence length alone, children's prose rates 'fairly easy' to 'fairly difficult' by Flesch's norms. When mean number of syllables per one hundred words is calculated, the Flesch norms of 'easy' or 'fairly easy' are satisfied. Again, these indices are surprisingly constant over the seventy-three-year period studied.

In the *St. Nicholas* material, the November and February issues, for all years, were consistently above May and August issues in Human Interest score. The relationship to Thanksgiving Day and to Valentine's Day, Lincoln's and Washington's birthdays is apparent. Probably the December issues would exceed even the Thanksgiving issues in this regard, although this hunch was not investigated.

Interesting as the results of word counts may be, such analyses miss completely the content of the material. 'Content analysis' has attracted the attention of journalists and social psychologists for some years. Of several techniques investigated, that proposed by Child and his associates[1] and applied by them to third-grade readers appeared to be most useful. The method consists in analyzing the behavior of principal characters in selected stories according both to the presumed motives of those characters and to the consequences of their behavior.

The psychological theory of behavior which lies behind this method is obvious. It is that children learn motives from observing the motives of persons they see or about whom they read. The assumption is made that the written page is an important source of conduct motivation in children, a position which has been and is implicit in a great deal of educational enterprise. Regardless of whether children's conduct is measurably affected by such material, one may analyze the material to discern the kinds of values set before children three-fourths of a century ago and more recently.

For this 'content analysis,' stories are selected according to the following criteria: (a) stories must contain action by animate characters, (b) they must not be essentially humor stories, for in such stories behavior and its consequences are frequently distorted for comic effect, and (c) they must not limit behavior descriptions to purely objective and external accounts—evaluation of behavior and its consequences must be implied (or at least the possibility of such evaluation must be implied). Judged by these criteria, the stories of *St. Nicholas* are excellently suited to analysis. They are full of people to whom many things happen; consequences of behavior are very apparent; it is easy to reach an evaluation of the consequences as desirable or undesirable for the character involved.

Stories which satisfy the above criteria are then analyzed in thema. A thema is a 'sequence of psychological events,' consisting of (1) a situation confronting a person or animate being, (2) the behavior, internal or external, with which the person responds, and (3) the consequences of the behavior as felt by the person or animate being himself. This third aspect of the thema, the present writers would restate as 'consequences of the behavior as felt by the person himself or as inferred to him by the reader.' In Child's work with materials prepared for third-grade children, consequences were usually verbalized. *St. Nicholas*, designed for more mature children, sometimes presents consequences in very subtle, inferential ways, almost beyond stated feelings at times.

Once thema are identified, they are classified in terms of an elaborate set of categories based by Child on Murray's[5] personality 'needs.' The writers used Child's categories and, with occasional simplifications, his definitions. The categories appear in Table III. Complete definitions of the categories are eliminated in the interest of brevity, but some examples follow (adapted from Child):

Activity.—Thema in which activity is engaged in for its own sake, and the emphasis in the story is on the value of the activity itself.

Achievement.—Incidents in which characters show either a generalized ambition to succeed or a desire for specific physical, intellectual, or social success.

Affiliation.—Thema including manifestations of friendliness and good will, and of the desire to do things in company with others. Often demonstrated through respect, deference, sympathy, generosity and helpfulness.

Nurturance.—Thema in which helpfulness and willingness to exert oneself to aid any creature that may be in need of assistance.

TABLE III.—THEMAS FOUND IN EACH BEHAVIORAL CATEGORY FOR EACH OF THE FOUR SAMPLED YEARS OF *ST. NICHOLAS*

Category	1873–4		1893–4		1913–4		1933–4	
	No.	Per cent	No.	Per cent	No.	Per cent	No.	Per cent
Objectless Behavior								
Activity	8	7.4	6	3.6	4	1.7	14	5.6
Passivity	1	.9	2	1.2	2	.8	0	0.0
Sentience	2	1.8	1	.6	0	0.0	0	0.0
Elation	2	1.8	2	1.2	3	1.3	2	.8
Total	13	11.9	11	6.6	9	3.8	16	6.4
Behavior Primarily in Relation to Things and Events								
Cognizance	1	.9	10	6.0	28	11.7	24	9.7
Achievement	16	14.7	23	14.0	37	15.5	34	13.8
Construction	1	.9	2	1.2	5	2.1	2	.8
Imaginality	0	0.0	6	3.6	0	0.0	1	.4
Acquisition	4	3.7	6	3.6	2	.8	5	2.0
Retention	2	1.8	1	.6	0	0.0	0	0.0
Order	2	1.8	1	.6	1	.4	1	.4
Total	26	23.8	49	29.6	73	30.5	67	27.1
Altruistic Social Behavior								
Affiliation	10	9.2	17	10.2	61	25.5	70	28.3
Nurturance	16	14.7	22	13.3	12	5.0	11	4.4
Succorance	6	5.5	6	3.6	9	3.8	7	2.8
Deference	5	4.6	11	6.6	4	1.7	2	.8
*Praying for others	1	.9	1	.6	0	0.0	0	0.0
*Appreciation	2	1.8	3	1.8	2	.8	0	0.0
Total	40	36.7	60	36.1	88	36.8	90	36.3
Egoistic Social Behavior								
Aggression	3	2.8	14	8.5	18	7.6	19	7.7
Dominance	8	7.3	7	4.2	11	4.6	10	4.0
Recognition	0	0.0	1	.6	2	.8	0	0.0
Autonomy	4	3.7	5	3.0	3	5.4	2	.8
Rejection	8	7.4	7	4.2	11	4.6	19	7.7
Total	23	21.2	34	20.5	45	23.0	50	20.2
Avoidance Behavior								
Harmavoidance	4	3.7	8	4.8	7	2.9	9	3.6
Blamavoidance	2	1.8	4	2.4	4	1.7	9	3.6
Infavoidance	1	.9	0	0.0	3	1.3	7	2.8
Total	7	6.4	12	7.2	14	5.9	25	10.0
Total of totals	109	100.0	166	100.0	229	100.0	248	100.0

* Categories added by the writers.

Aggression.—Thema featuring displays of temper and of flagrant disobedience; verbal forms of aggression such as criticisms and arguments; dominating others by bullying and teasing; physical attack and killing.

Harmavoidance.—Efforts to escape or avoid sources of danger.

The writers added two categories to those supplied by Child, which appeared necessary for the most accurate placement of thema. These are starred in Table III.

For the most part, the assignment of thema for categories was easy. When a single act might fulfill several purposes at the same time, and hence be assigned to multiple categories, the judgment was made in terms of the behavior which was predominant or most conspicuous. When consequences of behavior were classified under the headings 'reward,' 'punishment,' 'no consequence' (neither reward nor punishment), the permanent, last, or long-run consequences were considered, not the immediate consequences.

The sample of stories selected for analysis was drawn at twenty-year intervals, beginning with the year 1873–4. The first two stories of at least a page in length in each month's issue were used. There were in all ninety-six stories, twenty-four in each of four samples at intervals of twenty years. Serial stories were not included, and in only two instances was it necessary to include more than one story by the same author.

Results are summarized in Tables III to V. The first of these, Table III, shows the number of thema represented in each year's sample, and the proportions within each year represented in the several categories. The increase in gross number of thema observed in twenty-four stories reflects the increased length and especially the increased complexity of individual stories, which came somewhere around the turn of the century. Early issues contained many short stories. More recent decades have witnessed the introduction of mystery and adventure stories, where brother and sister hero-and-heroine combinations wander in and out of complex situations, exploring haunted houses, succoring destitute widows, and having exotic adventures in foreign lands. Yet the percentages vary little among the categories of thema, from sample to sample.

When the classes of categories are examined, the percentages appear very stable from period to period. There has been a decline in the amount of 'objectless behavior' (activity engaged in for its own sake, emphasis on emotion or feeling, etc.) and a corresponding increase in 'behavior primarily in relation to things and events,' but the difference between 1873 and 1933 falls short of the five per cent level of significance in both instances. Certainly, the emphasis on more concrete, activity-centered stories has not increased over the years to the extent we perhaps have believed.

The uniformly low percentages of avoidance behavior in all samples are of interest. Story characters have not and are not represented as seeking in their behavior to avoid physical harm, parental or social disapproval, fears, embarrassment, self-consciousness or shyness. In general, thema are distributed as follows: one-fifth to egoistic social behavior, one-third plus to altruistic behavior, and one-fourth to behavior primarily related to things and events, especially the gaining of knowledge and the desire for success. This is true both for the Seventies of the last century and the Thirties of the present.

The only differences which are statistically significant and which show a trend, occur within the category 'altruistic social behavior.' Here, there is a significant increase in behavior designated as 'affiliation-manifestations of friendliness and good will and of the desire to do things in the company of others; often demonstrated through respect, deference, sympathy, generosity, and helpfulness." This increase is paralleled by decreases in 'nurturance,' 'succorance,' and 'deference.' When extremes are compared, the decline in nurturance satisfies the one per cent level of significance. Nurturance is defined as "manifestations of helpfulness or willingness to exert oneself to aid any creature that might be in need of assistance." Succorance refers to "behavior directed at obtaining protection, sympathy, or assistance from another person," while deference indicates "behavior toward other persons involving (a) compliance with requests or commands, (b) politeness, (c) subservience, (d) respect for authority, and (e) imitation of other persons." In the last two categories, the trend is consistent but the shift is not statistically significant, at least in terms of samples of this size.

Table IV gives the number of thema in the more popular categories, all years combined, with the proportions in which the actor is rewarded, punished, or experiences no consequences of

TABLE IV.—CONSEQUENCES OF BEHAVIOR IN SELECTED CATEGORIES OF THEMAS, *ST. NICHOLAS* STORIES

Category	No. of Thema	Percentage in Which Behavior Was:		
		Rewarded	Punished	No Consequences
Objectless Behavior				
Activity	32	.35	.25	.39
Behavior Primarily in Relation to Things and Events				
Cognizance	60	.82	.05	.13
Achievement	109	.72	.22	.06
Construction	10	.50	.13	.37
Acquisition	17	.45	.33	.22
Altruistic Social Behavior				
Affiliation	158	.79	.03	.18
Nurturance	61	.71	.07	.22
Succorance	29	.75	.18	.07
Deference	22	.71	.10	.19
Egoistic Social Behavior				
Aggression	54	.44	.31	.25
Dominance	35	.39	.17	.44
Autonomy	24	.32	.33	.35
Rejection	45	.08	.11	.81
Avoidance Behavior				
Harmavoidance	27	.61	.12	.27
Blamavoidance	19	.15	.28	.57
Infavoidance	11	.21	.48	.31

his behavior. In general, altruistic social behavior is rewarded about three-fourths of the time. Egoistic social behavior is much less frequently rewarded; it is somewhat more likely to result in no consequences than to be punished. In behavior primarily related to things and events, the seeking of knowledge (cognizance) is rewarded in the great majority of instances; behavior denoting the desire for material possessions (acquisition), or the creation of a material or intellectual product for utilitarian or esthetic use (construction) is rewarded about half the time.

When the thema, distributed in the various categories by year samples, are broken down according to type of consequences, the number in any category becomes so small as to be relatively meaningless. The material, therefore, is not reproduced here. The totals, however, showed that for all categories combined there has been relatively little shift over the years in the tendency to reward or punish the consequences of behavior in stories. Consistently in samples taken at twenty-year intervals, behavior is rewarded in about three of five thema; it is punished in about one of six thema. For the remainder of the instances, no consequences follow from the behavior. An examination of the several categories of thema, such as objectless behavior, altruistic social behavior, egoistic social behavior, etc., shows that there has been no trend over the years to reward or to punish such classes of behavior in greater or less amount.

TABLE V.—CHARACTERISTICS PORTRAYED IN 24 STORIES SELECTED IN EACH OF FOUR YEARS, *ST. NICHOLAS* MAGAZINE

	1873–4		1893–4		1913–4		1933–4	
	N	Per cent	N	Per cent	N	Per cent	N	Per cent
Child	35	52.2	31	32.0	41	31.8	29	21.6
Adult	23	34.3	39	40.2	75	58.1	82	61.2
Animal	4	6.0	16	16.5	5	3.9	11	8.2
Fairy	0	0.0	5	5.2	3	2.3	2	1.5
Groups of the above	5	7.5	6	6.2	5	3.9	10	7.5
Total	67	100.0	97	100.1	129	100.0	134	100.0

Table V contains some additional data of interest. The total number of characters portrayed, per sample of twenty-four stories, has increased steadily. This is in keeping with the observation made earlier, that stories are longer and more complex now-a-days. The proportion of child characters has been cut in half over the years covered by this survey, and the proportion of adult characters has almost doubled.*

---

* These figures may, in part, reflect an increasing problem which the writers faced—that of determining the age of child characters. We took

It should be clear to the reader that this content survey covers only one source of children's literature, and a top-grade one at that. Yet, one impression stands out. Despite two world wars, an industrial revolution which has altered almost beyond belief the material comforts and conveniences of living, very marked alterations in habits of family living, major shifts in our economic ideologies, and tremendous advances in the format of children's reading materials themselves, the content of this sample of children's literature has altered very little indeed. To be sure, in this study we have established no absolute norm against which to measure change. Possibly, another set of content categories would give more evidence of change. Yet, the categories used appeared adequate to the writers; none disappeared, and no new ones were required for analysis of more recent materials.

The writers have carried from their study one or two impressions which cannot be substantiated from the tables, but which they feel inclined to report anyway. As stories become longer and more complex, there occurred a subtle shift in how child characters were portrayed. Children in stories of an earlier day were recognized as immature. Their interests, activities, preoccupations were not those of adults. They met failure and success, they were rewarded or punished, but as children, not as 'little adults.' Stories of a later era not only included fewer child characters but presented children more frequently as solving problems which baffled adults. While the industrious child of an earlier era was likely to win out over economic odds eventually, the boy detective of a later era was just as likely to solve a baffling mystery. The child character of an earlier day won by using materials at his disposal; the child character of a later period won by using, many times, the equipment and materials ordinarily available only to adults.

It has often been said that one major change in the thinking of adults in the last one hundred years has been to accord to children a life of their own, appropriate to recognized developmental levels. In one sense, however, the children's literature of an earlier day achieved this recognition somewhat more successfully than children's literature of a more modern time.

---

age sixteen as a division point, and when ages were not indicated, we made the best judgment possible in terms of the content. The trend is too marked, however, to be accounted for in terms of unreliability.

The writers would like to suggest that somewhat different ideological dimensions, developed around the hypothesis just advanced, would be worth studying in samples of children's literature. Religious stories and pamphlets for children should be included, likewise. Possibly, here evidence of greater adaptation might be found. As far as *St. Nicholas* magazine is concerned, it was surprisingly modern, even in its early days. Whether that be an accolade to *St. Nicholas'* early issues, or a condemnation of its later years, only further surveys of other materials for children can tell.

REFERENCES

1) Irvin L. Child, Elmer H. Potter and Estelle M. Levine. "Children's Textbooks and Personality Development." *Psychol. Monogr.*, 1946, 60, No. 3, pp. 1–54.

2) Rudolf Flesch. *The Art of Plain Talk.* New York: Harper and Brothers, 1946.

3) ——. *Marks of Readable Style: A Study in Adult Education.* New York: Bur. of Publ., Teachers Coll., Columbia Univ., 1943. (Centr. to Educ. No. 897).

4) ——. "A New Readability Yardstick," *J. appl. Psychol.*, 1948, 32, 221–233.

5) H. A. Murray, Jr. *et. al.*, *Exploration in Personality.* New York: Oxford University Press, 1938.

CORRECTIVE READING: A MEANS OF INCREASING SCHOLASTIC ATTAINMENT AT THE COLLEGE LEVEL

DOROTHY J. McGINNIS

Associate Director, Psycho-Educational Clinic
Western Michigan College

The purpose of this study is to describe and evaluate materials and procedures employed in doing corrective work in reading at the college level. Appraisal is made not only in terms of differences in initial and final scores on a reading test, but in point-hour-ratios as well. In making this evaluation one control and one experimental group have been equated, resulting group data have been compared statistically, a brief history of an individual in the Reading Laboratory has been presented, and tentative inferences have been made.

DESCRIPTION OF MATERIALS

(1) The primary source of materials used in the Reading Laboratory at Western Michigan College is the *Reading Manual and Workbook.*[1] This book contains twenty-two specific assignments, each of which has been designed to aid the student in developing a definite reading ability and skill. The reading assignments are related to actual classroom situations and the student is given an opportunity to apply the suggestions as he does his daily class preparation. Reading and studying are considered together. Assignments show how:

(a) To discover the nature and extent of the student's reading difficulties. (b) To aid in interpreting reading difficulties and in selecting the assignments upon which the student is to work. (c) To investigate the possibility of a visual defect which would interfere with reading performance. (d) To improve the student's ability to use the different parts of a book. (e) To develop skill in reading for a particular point and for general ideas. (f) To develop ability and skill in outlining. (g) To develop skill in reading a chapter effectively and well. (h) To develop skill in associating facts obtained from a textbook with facts already known. (i) To learn how to concentrate upon a reading activity. (j) To increase rate of reading without interfering with comprehension. (k) To develop skill in the use of the dictionary. (l) To increase reading vocabulary and to develop adequate skills in word study. (m) To develop skill in interpreting a short poem. (n) To develop skill in reading the newspaper effectively and well. (o) To develop ability and skill in problem-solving, especially in such areas as mathematics, chemistry, or physics. (p) To develop ability to read a table or graph. (q) To develop ability to read a map. (r) To develop skill in the use of the card catalogue. (s) To develop ability in using the *Readers' Guide*. (t) To develop skill in applying what is read to the solution of one's personal problems. (u) To develop skill in reading selectively and in preparing for examinations. (v) To appraise the change of reading skills and abilities acquired during the time between the initial and the final testing.

(2) *Orientation Inventory.*—This inventory, which is a part of *Reading Manual and Workbook*, has been designed to help the student recognize and pick out the reading abilities and skills which he does not possess and at the same time it furnishes information concerning his reading background and certain physical and emotional factors which may affect his reading performance.

(3) Equivalent forms of the Iowa Silent Reading Test[3] have been used. "The test is designed to cover a wide range of the skills known to be indispensable to effective reading of the work-study type. The test measures three broad general areas of silent reading abilities; namely, (a) Rate of Reading at a Controlled Level of Comprehension, (b) Comprehension of Words, Poetry, Sentences, Paragraphs, and Longer Articles, and (c) Ability to Use Skills Required in Locating Information."

(4) *Films for the Improvement of Reading.*[4]—These sixteen films, consisting of selections taken from a national journal, are intended to make each student read as swiftly as he can comprehend and no faster. This is done by projecting on the screen a printed page of film on which a broad band of light slides smoothly from left to right and from line to line. The student is not forced to stay within the band of more brightly lighted words but he is persuaded to attempt to follow this pacer. Comprehension check questions and transfer selections accompany the films.

PROCEDURE FOLLOWED IN THE READING LABORATORY

Persons who enter the Reading Laboratory do so because of their own initiative or as a result of advisement. No one is compelled to enroll. The greater proportion of individuals begin their work during the first week of the semester. Students, however, are permitted to enter any time if there is room in one of the sections. Each section is limited to twelve.

At the first meeting the members of the group are given the Orientation Inventory to complete. Next they are shown how to score this inventory and how to prepare a profile sheet which is of value to both the student and the instructor in interpreting factors underlying the reading difficulty.

During the next two meetings of the group Form Am of the Iowa Silent Reading Test is administered. The administration of this test requires approximately ten minutes of the third laboratory period. The remainder of the hour is spent in acquainting students with what has been accomplished by other people in the Reading Laboratory and an attempt is made to have them realize that they can improve their reading ability and at the same time facilitate their preparation for their regular college classes. The responsibility of learning to read is placed upon the student.

During the fourth meeting of the group the results of the Iowa Silent Reading Test and the Profile Sheet for Orientation Inventory are interpreted and each person is given an opportunity to ask questions concerning his performance. Class discussion is encouraged at this time. If physical or emotional abnormalities are indicated by the Orientation Inventory, the services of the Health Service or Psycho-Educational Clinic are requested.

The student is next asked to indicate on the Master Work Sheet the assignments upon which he needs to work in order to improve his reading ability. This Master Work Sheet is based upon the results of the Iowa Silent Reading Test and the Orientation Inventory.

Each person then completes Assignment III at which time he and the laboratory clinician investigate visual defects which may interfere with reading ability. If significant visual difficulties are found, he is referred to an ophthalmologist.

From this point on each individual is permitted to work upon the assignments indicated on his Master Work Sheet. He is 'on his own.' Seldom, if ever, do any two students have the same pattern of assignments. Each person is allowed to progress from assignment to assignment at his own rate and in accordance with his own plan. Some may spend days and even weeks upon one assignment while others may require much less time. The student, with the aid of the laboratory clinician, determines when he is ready for the next assignment indicated on his Master Work Sheet.

Individuals in the Reading Laboratory are urged to apply the suggestions made in the laboratory to their outside study. Many apply the various assignments in their class preparation of history and English, while others use their textbooks in biology and chemistry as a means of increasing their reading abilities and skills. Adjustment of materials to the reading abilities of the students is essential and frequently becomes a real problem for the instructor.

When the enrollee and the instructor agree that the purpose for which the student has entered the Laboratory has been accomplished, Form Bm of the Iowa Silent Reading Test is administered. Changes between initial and final test scores are interpreted and discussed. Many individuals continue two and even three semesters in the Laboratory.

PROCEDURE USED IN MAKING THE EVALUATION

1) Twenty college students enrolled in the Reading Laboratory were selected to make up the experimental group.

2) The Iowa Silent Reading Test, Form Am, was administered.

3) After a period of sixteen weeks an equivalent form of the Iowa Silent Reading Test was given to the twenty individuals enrolled in the Reading Laboratory.

4) Data resulting from the administration of the Iowa Silent Reading Test were tabulated and treated statistically.

5) One control and one experimental group were equated. In equating the groups the individuals were matched according to sex, status in college, scholastic aptitude and ability in reading. The control group consists of twenty persons selected from the student body at large who have not had work in the Reading Laboratory, whereas the experimental group consists of twenty individuals who have been actively engaged in the work of the Reading Laboratory. Each of the two groups consists of fifteen

men and five women. Eleven individuals making up each group are freshmen, five are sophomores, two are juniors and two are seniors. For both groups the median score on the Ohio State Psychological Examination is 49.5 and on Part III of the same measure 59.5. The mean number of hours carried by each group was 15.2. Inspection of courses suggest that reading requirements were approximately equal.

6) The point-hour-ratios for one semester of the control and the experimental groups were tabulated and treated statistically.

7) Data resulting from this study have been summarized and inferences have been tentatively set forth.

8) In this study no attempt has been made to determine the permanency of gain in reading ability, whether or not there is an increased achievement in other language skills, and whether or not gain in reading achievement is accompanied by constructive personality changes. Future investigations could well be directed to an appraisal of these factors.

TREATMENT OF DATA

Data resulting from this study have been tabulated as shown in Tables 1 and 2. In order to determine the significance of the difference between the means obtained in the initial and final testing, the single group method of obtaining the standard error of the difference between correlated means of small samples was used and '*t*' was found. In determining whether or not differences between the experimental and control groups were significant, the standard error of the difference between two uncorrelated means of small samples was found and '*t*' was calculated.[2]

BRIEF HISTORY OF AN INDIVIDUAL WORKING IN THE READING LABORATORY

Kenneth, who is eighteen years old, was graduated with a poor academic record from a large high school. Inability in reading was reported to be the chief cause of his poor school adjustment. However, corrective or remedial treatment in reading was not provided. After being refused permission to enter two midwestern colleges because of poor high-school grades, Kenneth sought the aid of the Psycho-Educational Clinic. The Wechsler-Bellevue Intelligence Scale revealed intelligence quotients of 119, 116, and 121 on the verbal, performance and full scales, respectively. Performance on the Iowa Silent Reading Test, Form Am, was equivalent to that of an individual eight months in the fifth grade (5.8). His percentiles on the Ohio State Psychological Examination and on the reading section of the same test were 23 and 20, respectively. The clinic suggested that Kenneth enter the Reading Laboratory at Western Michigan College where he also was permitted to enroll for fifteen hours of academic work. Assignments completed in the laboratory were those designed to develop skill in chapter reading, concentration, skimming, word study, use of the dictionary and skill in selecting essential facts for review. At the end of one semester his grade score on Form Bm of the Iowa Silent Reading Test was that of an individual five months in the tenth grade (10.5). This gain is statistically significant. In his academic work Kenneth earned three B's and three C's, which is equivalent to a point-hour-ratio of 1.4. He has enrolled for the second semester in the Reading Laboratory.

TABLE 1.—INITIAL AND FINAL IOWA SILENT READING TEST SCORES OF ALL STUDENTS IN THE EXPERIMENTAL GROUP

Median Standard Score	Initial	Final
200–204		5
195–199		3
190–194		
185–189	3	3
180–184	5	1
175–179	2	2
170–174		3
165–169	2	1
160–164	1	2
155–159	3	
150–154	3	
145–149		
140–144	1	
Total	20	20
Mean	169.5	185.25
Sigma	14.2	13.95

Standard error of the difference of the two means is 1.12, '*t*' = 14.06.

SUMMARY OF DATA

The distribution of initial and final scores of students enrolled in the Reading Laboratory on the Iowa Silent Reading Test is shown in Table 1. The correlation between equivalent forms was .94 and the degree of variability on both tests as indicated by sigmas of each distribution remained relatively constant.

In Table 2 a distribution is shown of the point-hour-ratios of students who have been enrolled in the Reading Laboratory and an equivalent group of individuals who have not done corrective work in reading. Variability as determined by the sigmas of each distribution is comparatively constant.

TABLE 2.—POINT-HOUR-RATIOS OF STUDENTS IN CONTROL AND EXPERIMENTAL GROUPS

Point-Hour-Ratio	Control	Experimental
3.00		1
2.75–2.99		
2.50–2.74		1
2.25–2.49		2
2.00–2.24	2	3
1.75–1.99	2	3
1.50–1.74	3	2
1.25–1.49	3	2
1.00–1.24	3	4
.75– .99		2
.50– .74	3	
.25– .49	3	
.00– .24	1	
Total	20	20
Mean	1.17	1.73
Sigma	.61	.605

Standard error of the difference of the two means is .196, 't' = 2.86.

Tables 1 and 2 show that persons enrolled in the Reading Laboratory made not only a statistically significant gain in reading ability as measured by initial and final tests, but also statistically higher point-hour-ratios than an equivalent group of students who were not enrolled in the laboratory. In both tables the differences between the means are significant at the one per cent level.

INFERENCES

1) The data of this study show quite conclusively that students who enrolled in the Reading Laboratory not only made a marked and statistically significant gain in reading as measured by an objective test, but also made statistically significant higher point-hour-ratios than an equivalent group of students who did not do corrective work in reading.

2) In consideration of the facts presented in this study it is reasonable to assume that the materials and procedures employed in the Reading Laboratory are of value in improving not only the reading ability of college students, but their point-hour-ratios as well. This, in part, may be due to the nature of the materials used in doing corrective work. The assignments have been directed primarily to the accomplishment of academic attainment and not merely to an increase in reading rate.

3) It is reasonable to assume that mechanical instruments and techniques used to facilitate adequate eye movement are useful adjuncts to the development of reading and study skills but that they cannot be regarded as *sine qua non.*

4) Motivation is a factor to be considered in the interpretation of data presented in this study for each student entered the laboratory voluntarily and not as a requirement.

5) In interpreting data resulting from this evaluation consideration should be given to the possibility that some gain in achievement may be due to the fact that students in the Reading Laboratory have been singled out and given individual consideration and attention.

REFERENCES

1) Homer L. J. Carter and Dorothy J. McGinnis, *Reading Manual and Workbook.* Prentice-Hall, Inc., 1949.

2) Henry E. Garrett, *Statistics in Psychology and Education.* Longmans, Green and Company, 1947.

3) H. A. Greene, A. N. Jorgensen and V. H. Kelley, *Iowa Silent Reading Test,* New Edition, Manual of Directions. World Book Company, 1943.

4) William G. Perry, Jr. and Charles P. Whitlock, *Instructor's Manual.* Harvard University Press, 1949.

AWARENESS OF VOCABULARY SIZE: ITS RELATION TO CLASS STANDING AND SEX DIFFERENCES*

JANICE LEVIN and CHALMERS L. STACEY
Syracuse University

INTRODUCTION

To discover the extent of the difference between the amount of knowledge that we credit to ourselves and the amount we actually possess is to discover our academic limitations and, not infrequently, the fact that we know more than we think we do in some respects and less in others.

Any college class taking an examination can be divided into three groups on the basis of expected and obtained grade on that test. One group will be made up of students whose expected grades approximate their obtained grades. The other two groups will be composed of students whose obtained grades are, respectively, higher and lower than expected. Is there any relationship between the students' standing in the course and the deviation between expected and obtained grades? Does the better student possess greater awareness of the degree of his mastery of the essentials of the course than the poorer student? How does the average student compare in this respect with the two extremes? Do sex differences play a rôle in such estimations?

That personality factors play an important part in the estimation of one's future performance has been shown in the many investigations of level of aspiration behavior. These factors have, of necessity, been neglected in this study.

Bear and Odbert's[1] study of the 'insight' displayed by college students in evaluating their knowledge of word meanings led them to conclude that "the average student's statement is a very uncertain guide to his performance on a test." Their two hundred eighty-two subjects indicated the words that they did not know in the material given them and were administered shortly afterward the vocabulary test from which the words were taken. The subjects missed sixty-four per cent of the words they said they did not know and forty-four per cent of the words they claimed to know.

---

* Presented at the meeting of Upper New York Psychologists at Syracuse University on November 5, 1949.

It is the aim of this paper to study awareness of vocabulary size in relation to class standing and sex differences.

DESIGN OF THE EXPERIMENT

The subjects of this study were one hundred forty-three sophomore students, sixty-seven males and seventy-six females, enrolled in an introductory psychology course. In respect to the study of class standing, three groups were formed on the basis of the sum of the raw scores on four one-hour examinations during the semester: lower twenty-five per cent, thirty-six subjects; middle fifty per cent, seventy-one subjects; upper twenty-five per cent, thirty-six subjects. In the testing which followed, these groups were compared with each other and with the entire group in respect to the measures being considered. With the exception of a female student in the lower twenty-five per cent group, all subjects took the tests relevant to this study. In the course of the study of sex differences, the male, female, and entire groups were compared with each other on the measures under consideration.

Development of the key.—Familiarity with words cannot justifiably be discussed in terms of a dichotomous relationship existing between knowing and not knowing the meanings of words. The continuum underlying word familiarity ranges from absolute unfamiliarity with the word to absolute certainty of its meaning. When subjects are forced to place words into 'know' and 'don't know' categories, there is no way of determining how familiar they are with the words they say they know or how unfamiliar with the words they claim not to know.

The familiarity of some words is beyond question. Some are more familiar in certain contexts than when by themselves or in other contexts. There are some words that we recognize and perhaps even use yet whose meanings are vague. Of the words whose meanings we do not know, there are some that we recognize and some that we cannot recall having seen or heard. These varying degrees of familiarity and strangeness of words, as discussed by Bear and Odbert,[1] served as the basis for the key which was used by the subjects for the classification of the words and phrases presented to them. The five items of the key were phrased as follows:

1) I know this word (or phrase). I know what it means no matter where it may appear.

2) This word (or phrase) is familiar to me. If it were used in a sentence, I would know what it means.

3) I recognize this word (or phrase). I am not sure that its use in a sentence would help me to know what it means.

4) I recognize this word (or phrase). I have no idea of its meaning.

5) I have never seen or heard this word (or phrase).

A Vocabulary Appraisal Sheet.—The Michigan Vocabulary Profile Test served as the basis for the preliminary data sheet which was called 'A Vocabulary Appraisal Sheet.' This vocabulary test has eight subtests covering the fields of study and interest of students. In each of the subtests are thirty items arranged in order of increasing difficulty. The items either define or describe one of four words or phrases given.

The last fifteen correct answers in seven subtests (the sports test was excluded) were listed on the appraisal sheet. Opposite each word or phrase were five boxes each of which represented one item of the key. The subjects were directed to familiarize themselves with the key items and the sample appraisal printed on the sheet and to express how well they felt they knew each word or phrase listed by placing an 'X' in the appropriate box next to each item.

The vocabulary test was administered four weeks after the subjects had filled out the appraisal sheet. Raw scores on each of the subtests and a composite raw score were obtained for each subject.

Combining key items.—On inspection of the item statements of the key, it will be noted that categories 1 and 2 indicate positive familiarity with the word; that is, within certain limits, the subject would be expected to know those words placed in these categories. Category 3 implies that the subject is uncertain of his ability to define the word and may be considered an 'uncertain' category. Items 4 and 5 are statements of negative familiarity: one would expect that the subject would define correctly only that proportion of words placed in these categories that can be predicted on a chance basis. In view of the combinative possibilities presented by the five-item key, three new categories were evolved, the positive, uncertain and negative.

The estimate scores are concrete demonstrations of awareness of vocabulary size. They represent the difference between estimation and performance. The positive estimate score is the difference between the number of words of those appraised which the subject claims to know and the number of words defined correctly on being tested. The negative estimate score is the number of words the subject says he does not know minus the number he fails to define on being tested. By definition, awareness of vocabulary size is a function of the estimate scores.

The positive estimate score was obtained by subtracting from the total number of words placed by the subject in categories 1 and 2, the total number defined correctly. A positive value of this measure indicates overestimation on the part of the subject of his ability to define the words presented. The negative estimate score is the sum of words placed in categories 4 and 5 minus the number missed. A positive value indicates that the subject has overestimated the number of words he would miss, or underestimated the size of his vocabulary in respect to these words. Positive and negative estimate scores were computed for each subject and averaged to yield group values.

STUDY I: AWARENESS OF VOCABULARY SIZE AND CLASS STANDING

AIMS OF THIS STUDY

Defining awareness of vocabulary size as a function of the deviation between what the subject says he knows and what he knows on being tested, we may say that the subject who possesses greater awareness of his vocabulary size will define correctly a greater percentage of the words he claims to know than the subject who is a poor judge. By the same token, one would expect him to overestimate and underestimate the number of words he knows or does not know to a smaller degree than the poorer judge.

With a clearer understanding of the tools and techniques used, it is now possible to state the null hypotheses, the testing of which may be considered a restatement of the aims of this study.

1) There is no difference between the three subgroups and entire group in respect to the mean percentage of correctly defined categorized words.

2) There is no difference between the subgroups and entire

group in tendency to overestimate or underestimate vocabulary size.

3) The correlations between the total raw scores on the vocabulary test and the semester raw scores for the subgroups and entire group are zero.

RESULTS AND DISCUSSION

In Table 1a are presented the mean percentages of correctly defined words placed in the categories of the five-item key by the subgroups and entire group. It will be noted that the upper group consistently defines a greater percentage of words correctly than the lower, middle, and entire groups. The middle group, with the exception of those words placed in category 2, defines a greater percentage of words than does the lower group.

TABLE 1A.—MEAN PERCENTAGES OF CORRECT DEFINITIONS OF WORDS PLACED IN EACH OF THE CATEGORIES OF THE FIVE-ITEM KEY BY THE SUBGROUPS AND ENTIRE GROUP

Category	Lower 25%	Middle 50%	Upper 25%	Entire group
1	57.2	59.8	67.6	61.5
2	46.0	41.8	52.1	46.6
3	39.7	40.8	43.1	41.2
4	29.0	32.3	41.5	34.3
5	25.0	25.7	28.1	26.3

Of the words placed in category 1, words which the subjects said they would know under all circumstances, the lower group missed forty-three per cent; the middle group, forty per cent; the upper group thirty-two per cent, and the entire group missed thirty-eight per cent. In category 2, the lower group failed to define fifty-four per cent of the words they said they would know in context; the middle group, fifty-eight per cent; the upper group, forty-eight per cent; and the entire group fifty-three per cent. The lower group knew forty per cent of the words they placed in category 3; the middle group, forty-one per cent; the upper group, forty-three per cent; and the entire group defined correctly forty-one per cent of the words of whose meanings they were unsure. Of the words placed in category 4, words which the subjects recognized but said they could not define, the percentage of correct definitions for the lower group was twenty-nine; for the middle group, thirty-two; upper group, forty-two;

and entire group thirty-four. A chance percentage of correct definitions was obtained for the words placed in category 5 by the subgroups and entire group.

Table 1b presents the mean percentages of correct definitions for the four groups in the categories of the three-item key. The upper group consistently defines a greater percentage of words correctly than the lower, middle, and entire groups. The middle group assumes an intermediate position in each category with respect to the other two subgroups.

TABLE 1B.—MEAN PERCENTAGES OF CORRECT DEFINITIONS OF WORDS PLACED IN EACH OF THE CATEGORIES OF THE THREE-ITEM KEY BY THE SUBGROUPS AND ENTIRE GROUP

Category	Lower 25%	Middle 50%	Upper 25%	Entire group
Positive	54.4	56.6	64.2	58.0
Uncertain	39.7	40.8	43.1	41.2
Negative	26.9	29.0	31.7	29.2

Critical ratios computed between the mean percentages of correct definitions in each category of the five- and three-item keys for the three subgroups and entire group yield no significant differences between them in respect to mean percentages of correctly defined categorized words.

The estimate scores of the subgroups and entire group are presented in Table 2a. As a whole, the groups overestimate the number of words they know in respect to both positive and negative estimate scores. The lower group presents evidence of greater overestimation of vocabulary size than does the middle group: this group, in turn, presents a greater deviation than the upper group. The negative estimate scores indicate that the

TABLE 2A.—MEAN ESTIMATE SCORES, STANDARD ERRORS, AND RANGES OF THE THREE SUBGROUPS AND ENTIRE GROUP

Estimate Scores	Measures	Lower 25%	Middle 50%	Upper 25%	Entire Group
Positive	Mean S.E.	18.1 1.7	17.4 1.2	9.8 1.7	15.7 .89
	Range	0 to 43	−6 to 38	−13 to 28	−13 to 43
Negative	Mean S.E.	−27.7 1.7	−29.0 1.3	−21.7 1.4	−26.8 .90
	Range	−52 to −10	−61 to −2	−40 to 0	−61 to 0

greatest underestimation of unknown words occurs in the middle group, while both the lower and upper groups underestimate less. On inspection of the ranges of the estimate scores for the subgroups, it will be noted that with rise in class standing the subjects tend to become more conservative in their estimations of vocabulary size.

In Table 2b are presented the critical ratios obtaining between the estimate scores of the subgroups and entire group. At the one per cent level, the differences between the upper group and the lower, middle, and entire groups are significant for both positive and negative estimate scores. In other words, while all groups tend to overestimate the number of words they knew, the upper group does so to a significantly smaller degree than the others.

TABLE 2B.—CRITICAL RATIOS OBTAINING BETWEEN THE MEAN ESTIMATE SCORES OF THE THREE SUBGROUPS AND ENTIRE GROUP

	Critical Ratios					
	L and M	L and U	L and E	M and U	M and E	U and E
Positive	.39	3.46	1.26	3.64	1.14	3.06
Negative	.60	2.73	.46	3.80	1.38	3.07

Correlations between semester raw scores and vocabulary test scores are presented in Table 3. The values of 't' were computed to test whether the obtained coefficients differed significantly from zero. The t-values yielded indicate that the correlations obtained for the lower and upper groups are not significantly different from zero, while the r-values of the middle and entire groups are significant at the two and one per cent levels.

TABLE 3.—CORRELATIONS OBTAINING BETWEEN SEMESTER RAW SCORES AND VOCABULARY TEST SCORES FOR THE SUBGROUPS AND ENTIRE GROUP

Group	r	S.E.	t
Lower 25%	.02	.17	.12
Middle 50%	.28	.11	2.42
Upper 25%	.07	.17	.41
Entire Gr.	.39	.07	5.03

CONCLUSIONS

1) The null hypothesis of equality of mean percentages of correctly defined words placed in the categories of the three- and five-item keys by the subgroups and entire group is not untenable with the data presented. One would therefore fail to reject the null hypothesis. However, despite the fact that the differences obtaining between the values of the measure for the different groups are not statistically significant, there is a consistent trend present in the direction of greater awareness of vocabulary size with rise in class standing.

2) The null hypothesis of equality of positive and negative estimate scores for the subgroups and entire group is rejected at the one per cent level. The upper group underestimates and overestimates to a significantly smaller degree than the lower, middle, and entire groups. Since awareness of vocabulary size has been defined as a function of the estimate scores, one may conclude that the upper group possesses a greater awareness of the size of its vocabulary than the other groups. The tendency of the middle group to manifest greater awareness of the size of its vocabulary than the lower group is present but not consistent.

3) On the basis of t-values presented, the null hypothesis of zero correlation between semester raw scores and vocabulary test scores is rejected at the two and the one per cent levels for the middle and entire groups. There is a trend in the direction of larger vocabulary size with rise in class standing.

4) In view of the data presented, one may conclude that as class standing rises, i.e., in rising from lower to middle to upper achievement groups in this group of subjects, greater awareness of vocabulary size is manifested.

STUDY II: AWARENESS OF VOCABULARY SIZE AND SEX DIFFERENCES

AIMS OF THIS STUDY

Awareness of vocabulary size, defined as a function of the deviation between estimated and obtained scores in a testing situation, has been found to be possessed by students to a remarkably small degree. It has been shown that, irrespective of class standing and vocabulary size, students in general tend to overestimate the number of words they know.

It is the aim of this paper to determine whether this poor judgment can be explained on the basis of sex differences in estimation of vocabulary size. The null hypotheses to be tested are the following:

1) There is no difference between males and females in mean percentages of correctly defined words placed in the categories of the three- and five-item keys of word familiarity.

2) There is no difference between males and females in respect to overestimation and underestimation of vocabulary size.

RESULTS AND DISCUSSION

Mean percentages of correctly defined categorized words.—In Table 4a are presented the mean percentages of correctly defined words placed in the categories of the five-item key by the male, female, and entire groups. The male group consistently defines a greater percentage of words than the female and entire groups. The entire group assumes an intermediate position in each category with respect to the male and female groups.

TABLE 4A.—MEAN PERCENTAGES OF CORRECT DEFINITIONS OF WORDS PLACED IN THE CATEGORIES OF THE FIVE-ITEM KEY BY THE MALE, FEMALE, AND ENTIRE GROUPS

Category	Male	Female	Entire Group
1	63.2	59.4	61.3
2	48.5	44.9	46.7
3	44.2	38.4	41.3
4	36.4	31.8	34.1
5	27.0	25.3	26.2

Of the words placed in category 1 of the five-item key, the males failed to define correctly thirty-seven per cent of the words, the females missed forty-one per cent, and the entire group missed thirty-nine per cent. The males failed to define fifty-two per cent of the words placed in category 2; the females, fifty-five per cent; and the entire group, fifty-three per cent. The males knew forty-four per cent of the words placed in category 3, the females knew thirty-eight per cent, and the entire group knew forty-one per cent. Of the words placed in category 4, the males knew thirty-six per cent; the females, thirty-two per cent; and the entire group, thirty-four per cent. A chance percentage of correct definitions of those words placed in category 5 was obtained for each of the groups.

Table 4b presents the mean percentages of correct definitions of words placed in the categories of the three-item key by the male, female, and entire groups. The males consistently define a greater percentage of words correctly than the entire group which, in turn, consistently defines a greater percentage of words than the female group.

TABLE 4B.—MEAN PERCENTAGES OF CORRECTLY DEFINED WORDS PLACED IN THE CATEGORIES OF THE THREE-ITEM KEY BY THE MALE, FEMALE, AND ENTIRE GROUPS

Category	Male	Female	Entire Group
Positive	59.6	56.5	58.0
Uncertain	44.2	38.4	41.3
Negative	31.0	27.6	29.2

Critical ratios computed between the three groups for each category of the three- and five-item keys yield no significant differences either between the sexes or between the entire group and the sex groups.

The estimate scores of the male, female, and entire groups are presented in Table 5. As a whole, the groups overestimate the number of words they know in respect to both positive and negative estimate scores. The female group consistently overestimates less than the male and entire groups. The male group consistently overestimates to a greater degree than the entire group. Critical ratios computed between the estimate scores of the three groups yield no significant differences in respect to sex differences in overestimation of vocabulary size.

TABLE 5.—MEAN ESTIMATE SCORES, STANDARD ERRORS, AND RANGES OF THE MALE, FEMALE, AND ENTIRE GROUPS

Score	Measure	Male	Female	Entire Group
Positive	Mean S.E.	16.4 1.2	15.0 1.3	15.7 .89
	Range	15 to 43	−13 to 38	−13 to 43
Negative	Mean S.E.	−27.4 1.3	−26.4 1.2	−26.8 .90
	Range	−3 to −61	0 to −51	0 to −61

CONCLUSIONS

1) The null hypothesis of equality of mean percentages of correctly defined words placed in the categories of the three- and five-item keys by the male, female, and entire groups is not untenable with the data presented. Although the males define correctly a greater percentage of categorized words than the female and entire groups, the differences observed are not statistically significant.

2) The null hypothesis of equality of estimate scores of male and female groups is not rejected. The males overestimate the number of words they know and underestimate the number they do not know to a greater, although not statistically significant, degree than the female and entire groups.

3) On the basis of the data presented, one must conclude that in this group of students, sex differences cannot account for the lack of awareness displayed by college students in respect to vocabulary size.

BIBLIOGRAPHY

1) R. N. Bear and H. S. Odbert. "Insight of older pupils into their knowledge of word meanings," *School Rev.*, 49, 1941, 754–765.

2) Edward W. Dolch. "Testing word difficulty," *Jour. Educ. Res.*, 26, 1932, 22–27.

3) Jeannette Elivian. "Word perception and word meaning in silent reading in the intermediate grades," *Educ.*, 59, 1938, 51–56.

4) R. A. Fisher. *The Design of Experiments.* 4th ed. New York, Hafner Publishing Co., 1947.

5) E. F. Lindquist. *Statistical Analysis in Educational Research.* Boston, Houghton Mifflin Co., 1940.

SOME COMMENTS CONCERNING THE RÔLE OF EDUCATION IN THE "CREATION OF CREATION"[1]

EDITH A. WEISSKOPF

Department of Psychology
Purdue University

Most people would agree that it is one of the main aims of education to encourage the development of creative abilities. Therefore, educators are forced to concern themselves with the following question: What are the conditions which enhance the formation of creative thought? The literature in psychology and related fields contains a considerable amount of information about creative thought and the conditions of its growth. This information is based on general impression, on the analysis of biographical data, on introspective statements of creative thinkers, or on experimental research. In spite of the diversity of its origin the literature shows a considerable amount of agreement as to some of the conditions enhancing creativity. Thus, one would expect educators to use techniques in their intellectual work similar to the techniques which are thought to create a fertile soil for the growth of new ideas.

Strangely enough this is not often the case. It is the objective of this paper to discuss some specific aspects of the well-known fact that the mental attitude during intellectual work encouraged by our schools is in many respects diametrically opposed to the mental attitude which is thought to stimulate creativity.[7,18]

Most educators in their oral or written admonitions to their pupils stress the importance of industry. Yet, if we scan through the literature on creativity, we find statements such as the following: " . . . though without industry great intellectual work can not be done, yet mere industry may prevent creation"[18] (p. 149). Educators emphatically stress the first part of this statement when teaching students, but are, with few exceptions (e.g.[10]), reluctant even to mention the second part. The constant activity enforced by many educators does not give young people the leisure which is an essential prerequisite for intellectual or artistic creation.[7,18]

---

[1] A paper presented at the Annual Meeting of the American Psychological Association, September, 1950.

We teach our pupils the importance of regular study habits. Why do we ignore statements such as the following? "The advantage of regular habits during the student years is great; but the optimum point at which the curve of that advantage cuts the curve of the advantage of fresh initiative is different for those whose professional work will be intellectual origination, and for those of different powers and aims"[18] (p. 297). If pupils would use the explicit admonitions received from their teachers as their only source of knowledge on study techniques, it would never occur to them that regular study habits could be anything but advantageous.

We try to instill a critical attitude towards intellectual matters in our students, but Hutchinson quotes a scientist saying: "It is surprising how fertile the mind becomes when not interrupted, or restricted by criticism in the free and often fantastic expression of its ideas"[6] (p. 32).

Our school system stresses the importance of intellectual objectivity and the liabilities of one-sidedness. Thus, " . . . students often start their careers with some intense enthusiasm, then learn in college how narrow and 'immature' it is, how much can be said on the other side, and how many other good ways of doing things exist. They gradually acquire a spirit of mellow tolerance and scholarly insight, but lose the creative impulse"[11] (p. 301).

It is challenging to examine in this connection suggestions made by psychiatrists or clinical psychologists on the mental attitude which leads to optimal results in producing insight and understanding with regard to their patients' problems. Freud, for example, makes the following recommendations on treatment techniques: "All conscious exertion is to be withheld from the capacity for attention . . . "[2] (p. 324). "One has simply to listen and not to trouble to keep in mind anything in particular" (p. 325). "It is not a good thing to formulate a case scientifically while treatment is proceeding, to reconstruct its development, anticipate its progress, and take notes from time to time of the condition at the moment, as scientific interest would require" (p. 326–7). " . . . the most successful cases are those in which one proceeds, as it were, aimlessly, and allows oneself to be overtaken by any surprises, always presenting to them an open mind, free from expectation" (p. 327). Reik also emphasizes the liabilities of continuous sharp attention and the benefits of inattentive states: " . . . training in attention may succeed "too well," so that the child is shut off from a whole flood of freely emerging associations and ideas"[17] (p. 165). "The concentration of attention involves a setting up of inhibitions, so that keen observation of particular things corresponds to the ignoring of others. Withdrawal of certain contents from consciousness is no less a part of attention than the appearance of others at the center. It is open to questions which part is the more significant to the essence and action of attention, the illumination of particular objects, the bestowal of intensive interest upon them, or its diversion away from others" (p. 164). "Not until I cast off the customary restraint of voluntary attention was I able to get hold of the hidden psychical data. Not until I left the firm and broad high road did I reach the goal along side paths. The secret meaning escaped my conscious, active attention, and was not found until I had become 'inattentive' in the popular meaning of the word, that is, until I gave myself up to unconscious ideas of the goal" (p. 169).

If a student would show a mental attitude towards his school work similar to the one advocated by these clinicians, he would elicit the serious criticism of his teachers. The objection may be raised that intellectual work of the academic type is basically different from the achieving of insight into the personality problems of a neurotic patient, and that it, therefore, requires a different mental attitude. Quite true. Yet the literature on creative thought suggests that Freud's admonitions may prove useful in any situation where insight is required. Moreover, Maier and Luchins show that the arriving at solutions of various kinds of problems is facilitated by such 'loosening' instructions as "Do not be a creature of habit and stay in a rut," "Keep your mind open for new meanings," "The solution pattern appears suddenly. You cannot force it"[9] (p. 147), or "Don't be blind"[8] (p. 2).

Thus, educators may find it profitable to advocate a passive, inattentive attitude to their students not as a permanent condition but as a stimulating technique during specific phases of their intellectual work. As Bateson states " . . . the advances in scientific thought come from a *combination of loose and strict thinking*, and this combination is the most precious tool of science"[1] (p. 55).

At this point I expect some readers to voice objections. "Industry, regular study habits, a critical, controlled attitude, are necessary attributes of successful intellectual work," they will say. "The creative individuals who found that relaxation, irregular habits, and temporary abandonment of criticism and control enhance creativity have spent years of strenuous preparation during which they exercised all the virtues advocated by traditional educators." Quite true. Yet the question remains unanswered why, among the four stages of the creative process, namely, preparation, incubation, illumination, and verification[3–7,12–16,18] we prepare children for the first and last stage only, and completely ignore the other two stages.[18] Only one possible cause of this educational bias will be suggested in this paper.

The process of insight and illumination appears to be directed by unconscious forces to a higher degree than the process of preparation and verification. Just as lightning represents the sudden explosive merging of positive and negative electricity, intellectual insight may represent the sudden merging of conscious content accumulated during the stage of preparation with unconscious, repressed material.[5,6] Thus, it may be the denial of this repressed material which causes educators to deny the process of illumination. In other words, we do not want to accept the fact that intellectual activity takes part of its energy supply from the big reservoir of unacceptable impulses called the Id, that the material carefully and rationally accumulated during the period of preparation has to become imbued with Id-impulses in order to be brought to life. We falsely admonish our students to exert control in order to keep these impulses repressed instead of teaching them to make their impulses subservient to intellectual activity.

Just as we often deceive small children about the nature of biological creation, we deceive older children by pretending that intellectual creation is unemotional. Thus, by de-emotionalizing intellectuality and by keeping the 'sweet secret' of intellectual creation from youth, we suffocate intellectual creativity.

BIBLIOGRAPHY

1) G. Bateson. "Experiments in thinking about observed ethnological material," *Phil. Sci.*, 1941, 8, 53–68.

2) S. Freud. *Collected Papers*, London: Hogarth Press, 1949. Volume II.

3) H. von Helmholtz. *Vorträge und Reden*, Braunschweig: F. Vierweg & Sohn, 1903.

4) E. D. Hutchinson. "Materials for the study of creative thinking," *Psych. Bull.*, 1931, 28, 392–410.

5) ———. "Varieties of insight in humans," *Psychiatry*, 1939, 2, 323–332.

6) ———. "The nature of insight," *Psychiatry*, 1941, 4, 31–43.

7) ———. "The phenomenon of insight in relation to education," Psychiatry, 1942, 5, 499–507.

8) A. S. Luchins. "Mechanization in problem solving. The effect of *Einstellung*," *Psychol. Monogr.*, 1942, 54, No. 248.

9) N. R. F. Maier. "An aspect of human reasoning," *British J. Psychol.*, 1933, 24, 144–155.

10) R. L. Munroe. *Teaching the Individual*, New York: Columbia University Press, 1942.

11) T. Munro. "Creative ability in art and its educational fostering," *Yearb. Nat. Soc. Stud. Educ.*, 1941, 40, 289–322.

12) C. Patrick. "Creative thought in poets," *Arch. Psychol.*, 1935, 26, No. 178.

13) ———. "Creative thought in artists," *J. Psychol.*, 1937, 4, 35–73.

14) ———. "Scientific thought," *J. of Psychol.*, 1938, 5, 55–83.

15) ———. "Whole and part relationship in creative thought," *Am. J. Psychol.*, 1941, 54, 128–131.

16) H. Poincaré. *Science et Méthode.* Paris: Flammarion, 1918.

17) Th. Reik. *Listening with the Third Ear.* New York: Farrer, Straus, and Co., 1949.

18) G. Wallas. *The Art of Thought.* New York: Harcourt Brace and Co., 1926.

CELIA B. STENDLER. *Children of Brasstown.* Urbana, Illinois: Bureau of Research and Service of the College of Education, University of Illinois, 1949, pp. 103.

This monograph is primarily concerned with the awareness of 'social class' and class symbols of first-, fourth-, sixth- and eighth-grade children in one school in a New England industrial town of 15,000 inhabitants.

The author identified four classes in the community 'by interviews with townspeople.' Unfortunately, we are not told how many people were interviewed, how they were selected, nor which 'classes' they represent. The one hundred seven children in the study were then rated by five adult judges as belonging to the upper middle, 'white collar,' or working class. An upper class category was not used by the judges since it was presumed that upper class children did not attend public schools. There seems to be a class bias in the ratings since none of the judges represented the working class. Further, in discussing the 'social classes' the author sometimes refers to them as socio-economic categories and at other times as prestige categories. It is unfortunate that more rigorous procedures were not used in stratifying the community and in class placement of the children, for the remainder of the research is carefully designed.

The children were interviewed in regard to their out-of-school activities and their friendship choices. They were asked to rate four sets of pictures representing class differentials in jobs, houses, clothing and recreation. In addition, they rated their own and classmates' 'social class' position and answered a 'guess who' test. Some of the major substantive conclusions were: (1) the higher the grade in general the greater the awareness of class symbols; (2) in the lower grades, boys are more discerning in awareness of class symbols and in the higher grades girls seem to be more discerning; (3) the higher one's class placement, the more conscious one is of class symbols; (4) out-of-school activities such as work activities and music and dancing lessons are conditioned by class placement of parents; (5) in-school choices are more democratic than out-school choices in friendship selections; (6) the higher the grade, the greater the awareness of class position of schoolmates and the greater the awareness of class as revealed in future occupational choice. Miss Stendler also offers certain generalizations applying to all class levels and sets up four levels of awareness of class symbols.

In the final chapter the author discusses the implications of her findings. She raises some trenchant questions regarding the unrealism of typical social studies programs in the primary grades and the proposed solutions of Allison Davis and others to the lower class child's anxiety over loss of prestige. Educational psychologists and school administrators will find this chapter exceedingly provocative.

This reviewer was somewhat dissatisfied with the omission of supporting tables and statistical analyses for certain conclusions. One would like to be able to judge for oneself, for example, the significance of the analysis of variance tests rather than accept the author's judgments without supporting data that certain relationships are statistically significant (p. 53). NEAL GROSS

University of Minnesota

E. G. WILLIAMS, Editor. *Trends in Student Personnel Work.* Minneapolis: University of Minnesota Press, 1949, pp. 417.

This book presents forty-three papers by nearly as many authorities, read at the University of Minnesota in connection with a three-day meeting of "associates in personnel work" from various institutions of higher education. The papers present (1) an overview of student personnel work during the last quarter century, (2) some of the difficulties encountered along the way, and (3) points up current best practices, covering research on methods, tests, counseling techniques and therapeutic procedures. This book can be read with interest and profit both by students and workers. "The papers outline the growing edge of student personnel work." It should be particularly helpful to anyone—student or professor—wishing to gain acquaintance with the field both as to present status and future promise.

The broad scope of the papers may be seen from the following listing of the main divisions of the book (contributors to each division are also here listed): Part I, The Rôle of Personnel Work in Higher Education (M. M. Willey, W. H. Cowley, M. S. MacLean and H. Y. McClusky). Part II, Developments in Tests and Testing (J. G. Darley and D. B. Stuit). Part III, Vocational Counseling and Occupational Adjustment (D. G. Paterson and C. L. Shartle). Part IV, Counseling Methods and Points of View (E. S. Bordin, E. H. Porter, Jr.). Part V, Mental Hygiene (R. G. Hinckley, R. F. Berdie and Mrs. Anne Fenlason). Part VI, Problems of Student Veterans (R. J. Strom and S. J. Wenberg). Part VII, Special Personnel Services (F. G. Moore, Mrs. Dorothy Dyer, W. C. Stevens, Mrs. Margaret Aldrich, J. D. Foley, Bryng Bryngelson, G. B. Risty, Ruth E. Boynton). Part VIII, The Rôle of Student Housing in Education (B. J. Borreson, A. H. Drammen). Part IX, Personnel Work as a Profession (Esther Lloyd-Jones, C. G. Wrenn and J. G. Darley). Part X, Counseling by Faculty Members (D. D. Feder, W. M. Gilbert, D. J. Shank and J. L. Bergstresser). Part XI, Religious Programs and Student Personnel Work (E. G. Williamson, Rev. James Boren, Rabbi Norman Frimer, Father Leonard Cowley). Part XII, Social Adjustment through Activities (W. H. Cowley, R. L. Sutherland). Part XIII, Personnel Problems of Women (Kate Hevner Mueller, Esther Lloyd-Jones). Part XIV, Testing and the Curriculum (C. P. Turner, R. W. Tyler).

J. B. Stroud

State University of Iowa

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 42 October, 1951 Number 6

COMMUNITY VARIABLES AS PREDICTORS OF INTELLIGENCE AND ACADEMIC ACHIEVEMENT

ROBERT L. THORNDIKE

Teachers College, Columbia University

INTRODUCTION

In 1939 E. L. Thorndike[3] reported the results of a detailed statistical investigation of the facts about American cities which serve to predict the extent to which a city provides a good life for its inhabitants. The criterion of the good life in this research was a rationally weighted composite of measures of the health, education, comfort, etc. of the population. He studied the relationship of the index of 'general goodness' to an index of community wealth and one representing the quality of the population. On the basis of the correlation between these indices, he concluded that 'general goodness' was more directly a function of the qualities of the people in the community than it was of community wealth, and that personal qualities and wealth in combination accounted for most of the predictable 'goodness' of the community.

Even before this time Maller,[2] studying districts in New York City, had found a substantial relationship to exist between school acceleration and community variables. He reported a multiple correlation of .81 between a composite of six economic and health variables and the index of school acceleration.

Thorndike and Woodyard[4] followed up Thorndike's original study with a project to test the children in a sample of cities and studied the relationship between community variables and performance on a standardized intelligence test. The National Intelligence Test was administered to all sixth-grade children in thirty small cities. On the basis of these data, estimates were made of the correlation between N.I.T. score and the indices of general goodness, personal qualities of the population and per capita income. It was estimated that the correlations were, respectively, .86, .82, and .78.

In none of these early studies were objective measures of intelligence or achievement available for any substantial number of communities. The gathering of test data on such a large scale merely for a research project would, of course, have been prohibitively expensive. It is only as data have become available from large scale testing or test standardization projects that studies of the relationship of community variables to measured intelligence or achievement have become possible. The present analysis is based upon the testing carried out in 1945 to standardize revised forms of the Metropolitan Achievement Test. The study of relationships between community characteristics and the measured intelligence and school achievement of the children in the community seemed both of theoretical interest in describing some of the sociological concomitants of ability and achievement, and of possible practical value in guiding the selection of communities for possible future enterprises of test standardization.

Since the present study was undertaken, a report has been published by Davenport and Remmers[1] describing relationships between social and economic variables for states and the average performance of young men from those states on the A-12 and V-12 qualifying examination. The group taking the test were young men in high school or college trying to qualify for this World War II educational program, but the exact nature of the group, or the degree to which it represented a select fraction of the population of that age in the different states can hardly be determined. Davenport and Remmers studied twelve predictors in relation to the test scores, and found substantial correlations for ten of them. The highest correlations were:

Number of telephones per 1000 population	.83
Per capita income	.81
Value of school property	.76

On the basis of a factor analysis of the test scores and the predictor variables, the authors conclude that an economic factor accounts for the bulk of the variation in test achievement. Certainly, the most predictive variables have an economic flavor, though the extent to which the economic component is cause and the extent to which it is symptom of underlying population characteristics can hardly be determined from a statistical analysis of this limited number of variables.

SOURCES OF DATA

Data required for the present study were of two types. In the first place, it was necessary to have measures of the level of intellectual ability and scholastic achievement for each community. These measures were obtained from the nation-wide testing that was carried out by the World Book Company as part of the program of establishing norms for the revised forms of the Metropolitan Achievement Test. In the second place, it was necessary to have a variety of educational, economic and sociological facts about each community for which test data were available. These were obtained from the reports of the 1940 census. The nature of these data and the way in which they were prepared for analysis will be discussed in some detail.

As indicated, intelligence and achievement test data were obtained from the standardization testing to provide norms for the revised form of the Metropolitan Achievement Test. These were supplied to the author by the World Book Company.* The standardization testing was carried out in October, 1945. The selection of communities in which to test was necessarily based in part on the willingness of the community to coöperate. However, through its regional agents the World Book Company was able to obtain communities spread over the whole United States, and test returns were received from at least one community in each state. Communities ranged in size from New York City, Oakland, Providence and Atlanta, where testing was carried out in a selected sample of schools, down to many communities with populations of under 2500. In addition, a number of county school systems were included, thus giving coverage of rural as well as urban areas. In all, something in the neighborhood of 500,000 children were tested, spread over some three hundred administrative school units.

* I wish to acknowledge my debt to the World Book Company for these materials, and particularly to Dr. Walter Durost, at that time Director of the Division of Test Research and Service, who called my attention to them and went to some pains to make them available to me.

For a school system to be accepted as a participant in the standardization program, it had to undertake to test all the children in at least three adjacent school grades. Many, however, tested over a wider grade span, and some tested in all grades from II through IX. The only exception to complete testing was in a few of the largest cities. In these cities, testing was carried out in certain selected schools, chosen by an administrator of the system as being representative of the schools of the community as a whole. All testing was carried out at nearly the same time of year, i.e., during October. Original scoring of the test blanks was carried out by the coöperating school systems. However, approximately a twenty per cent sample of test papers was returned to the World Book Company for check scoring. A sample was selected at random from all the papers in a given class, following procedures specified by the publisher. The papers which were returned to the publisher and rescored were the ones used for establishing test norms, and are the ones which provide the data for this study. In specifying the sample to be returned to them, the publishers called for a minimum of six papers from each class, no matter how small the class. With small classes the per cent returned was well over twenty, but with larger classes the returns were held to twenty per cent.

The testing program involved giving each class an appropriate form of the Metropolitan Achievement Test and an appropriate level of the Pintner Intelligence Tests, Verbal Series. The forms which were given to the different grades are as follows:

Grade	Intelligence Test	Achievement Battery
2	Pintner-Cunningham Primary	Primary I
3	Pintner-Durost (Picture Scale)	Primary II
4	Pintner-Durost (Reading Scale)	Elementary
5	Pintner Intermediate	Elementary
6	Pintner Intermediate	Intermediate
7	Pintner Intermediate	Intermediate
8	Pintner Intermediate	Advanced
9	Pintner Intermediate	Advanced

At the time the analyses reported in this study were undertaken, all tests had been scored and all raw scores had been expressed as derived or norm scores. Since the concern in this study was not to be with individual children, but rather with individual communities, data made available for an analysis were averages. Average IQ was reported in the case of the intelligence test and average standard score in the case of each subtest of the achievement battery for each grade level in each community.

Given data as indicated above for each separate grade group, it was necessary to consolidate them into a single intelligence index and a single achievement index to represent a given community. In the case of the intelligence test, this was very simple. Since the type of IQ reported for the Pintner tests is essentially a standard score, with approximately the same mean and standard deviation at different age levels, it was possible to combine data from different grades directly into a single average IQ. No attempt was made to take account of differences in the number of cases in different grade groups from a single community. These tended to be small in any case, and need not have corresponded too closely to total numbers in the different grades in the community. There seemed little reason to think that they would give a better index for the community than a single unweighted average, and so the unweighted average of the separate grade averages was used. This provided a single index of level of intelligence for the children of the community.

Obtaining a single achievement index for each community was somewhat more indirect and laborious. The grade averages which were available were for each subtest separately, and they were expressed in terms of a single scale of converted scores. The publishers of the Metropolitan Achievement Test have provided conversion tables for translating raw scores on each subtest into standard scores which are anchored to the sixth-grade group. The sixth-grade mean is defined as 200 and the standard deviation is 20. By anchoring all levels of the test for all grades to this reference point, they have tried to facilitate comparisons between different forms and different subtests. The average standard score for a grade group is not directly interpretable as a measure of degree of excellence for the group, or of its deviation from all groups of the same grade. Fortunately, the publishers had also developed percentile norms for each grade, by means of which it was possible to translate a given standard score into a percentile value for that grade. If the assumption of a normal distribution is made, these percentiles can in turn be translated into normalized standard score values expressing in more nearly equal units the degree of excellence of a particular score relative to its particular grade group. Standard scores with a mean of 50 and a standard deviation of 20 were used. Thus, a particular fourth-grade class group is found to have an average standard score on the arithmetic problems test of 170. This falls at the 70th percentile for fourth-grade children, and becomes a scale score of 60. The scale score designates the degree of excellence of that fourth-grade class as far as the measure of arithmetical reasoning is concerned.

The first step, then, was to translate the score of each grade group on each subtest into the corresponding scale score, representing degree of excellence based on the performance of all children of that grade level. The next problem was what to do with the results of the several subtests, numbering from four to ten depending upon the grade level. One possible procedure would have been to keep the subtests separate and to study the background factors as predictors of success in each separate area of achievement. This possibility was rejected for the initial analysis. It was felt desirable first of all to get a general picture of the factors related to achievement. The scale scores from the separate subtests were, therefore, combined into a single score.

In combining the subtests, the problem was encountered of how to weight the different content areas. This was brought to the fore by the fact that in certain areas, notably arithmetic and reading, there were more than one subtest. Also certain areas appeared for all grades while others, such as science and history, were included only in the test for the higher grades. The procedure which was finally adopted was to weight each subtest equally in computing an average scale score for each grade group, and then to weight each grade group equally in computing an average scale score for each community. Thus equal weight was given to a second-grade group and to an eighth-grade group, though the average was based on four subtests in one case and ten in the other. The final decision was based on the premises (a) that the allocation of number of subtests to an area at a given grade level represented the best judgment of the makers of the test and might well be followed and (b) that each grade's performance has equal significance in appraising the performance of the community as a whole. The net effect of this decision was to give relatively heavy weight to reading and arithmetic and relatively light weight to the content areas in the over-all score for a community.

In summary, then, the single achievement index was an average scale score. It was obtained by averaging the average scale scores for the separate grades in the community, weighting each grade equally. The grade average scale score was obtained by averaging the grade normalized scale scores for the several subtests. The subtest scores were obtained by using the table of percentile norms which related the universal standard scores for the subtest to percentiles for the particular grade.

After the operations which have been described had been carried out there were available for each community an average IQ and an average achievement scale score. With respect to the achievement scores, however, one additional problem presented itself. It seemed altogether likely that different communities differed considerably as to the age of the pupils in a given grade. Since achievement should be expected to be a function of age as well as of grade reached, it was necessary to take account of this fact. Accordingly, tabulations were made of the average age of pupils in each grade for each community. Distributions were prepared for all communities reporting data for a given grade, and the mean age and standard deviation of ages was determined for that grade. The values for the individual communities were converted into standard scores. The standard scores for the several grades in a community were averaged to yield an average age index for the community. This age index made it possible to hold age constant when reporting correlations with the achievement index.

Facts about the communities included in the study were taken from the reports of the 1940 census. The analysis was restricted to facts available in the 1940 census because of their availability in standard form for all communities of a given size and type throughout the country. Availability was of some importance to the experimenter in carrying out this study. It would be even more important in any use of the results of this study by another worker in order to select communities for norming a new test or in order to interpret the achievement of a particular community in which he was interested. Representative data of a variety of types were chosen, covering aspects of population structure, educational level, occupational distribution, housing conditions, and economic level. The choice of specific variables was admittedly somewhat arbitrary. It was limited to a considerable extent by what was available in published census reports. Within those limits the further choice represented the judgment of the experimenter as to which seemed meaningful and fruitful to explore. There are many other variables which might have been included and which it may prove fruitful to explore in a later study.

The variables which were studied as predictors of intelligence and achievement are listed below. Each variable is defined sufficiently precisely for it to be reproduced by the reader.

PREDICTOR VARIABLES

Population. (Variable 1) Population in thousands, as given in the 1940 census.

Per cent Native-born White. (Variable 2) This variable was obtained by dividing the number of native-born white by the total population.

Per cent Under 5 *Years Old.* (Variable 3) Census figures for the number under five years were divided by the total population.

Per cent Over 21 *Years Old.* (Variable 4) Census figures for the number over twenty-one were divided by the number in the total population.

Median Grade Reached by Adult Population. (Variable 5) This figure, taken directly from census reports, indicates the median school grade reached by adults twenty-five years old and over.

Per cent of Illiteracy in the Adult Population. (Variable 6) The number 25 years and over having four years schooling or less is divided by the total number 25 and over.

Per cent of Adults Who Are High-School Graduates. (Variable 7) Computed as for Variable 6, using those reporting twelve or more years of schooling.

Per cent of Males Employed. (Variable 8) Obtained by dividing the number of males fourteen and over gainfully employed in other than public works (WPA, etc.) by the total number of males fourteen and over.

Per cent of Females Employed. (Variable 9) A similar figure for females fourteen and over, computed in the same way as Variable 8.

Persons per 1000 Engaged in a Profession. (Variable 10) The number of men and women reported as working at a profession was divided by the total population of the community.

Male Domestic Servants per 1000 Population. (Variable 11) The number of males reporting their occupation as domestic service was divided by the total population of the community.

Female Domestic Servants per 1000 Population. (Variable 12) Computed in the same way as Variable 11.

Male Laborers per 1000 Population. (Variable 13) Computed in the same way as Variable 11.

Per cent of Homes Owner-Occupied. (Variable 14) The number of owner-occupied residential units was divided by the total number of residential units in the community.

Per cent of Homes Tenant-Occupied. (Variable 15) Computed in the same manner as Variable 14.

Per cent of Homes Overcrowded. (Variable 16) The number of residence units occupied by more than 1.5 persons per room was divided by the total number of residence units reporting.

Per cent of Homes Needing Major Repair. (Variable 17) The number of residential units needing major repair was divided by the total number which were reported on with respect to this item.

Per cent of Homes Lacking Private Bath. (Variable 18) Computed in the same manner as Variable 17.

Per cent of Owner-Occupied Homes Mortgaged. (Variable 19) The total number of owners reporting mortgages on their homes was divided by the total number responding to the question.

Median Home Rental. (Variable 20) Taken directly from census figures.

Per Capita Retail Sales. (Variable 21) The total of retail sales reported by the stores of the community was divided by the total population.

Per Capita Food Store Sales. (Variable 22) The total retail sales reported by food stores was divided by the total population.

Per Capita Retail Sales for County. (Variable 23) Computed in same way as Variable 21, but based upon the complete county in which the community was located.

Average Pay in Retail Stores. (Variable 24) Total pay to employees in retail trade was divided by the number of such employees reported for the community.

Data for the twenty-four predictor variables and the three community indices were punched on IBM cards and the complete matrix of intercorrelations was computed. These correlations were based on records for one hundred fifty-four towns and cities. Data were also available for twenty-eight county units, and these were used to 'cross validate' the values obtained for the urban units. The correlations among the twenty-four predictor variables are shown in Table I. The correlations of the predictors with the intelligence index, the age index, the achievement index, and the achievement index with age held constant are shown in Table II. This table also gives the mean and standard deviation of each variable and the correlations among the three community indices.

Interest focuses first of all on the correlations of the variables with the intelligence index. Inspection of the appropriate column of Table II shows a considerable range of values, with a number that are statistically significant. With one hundred fifty-four cases a correlation must exceed .20 to be significant at the one per cent level, and eleven of the correlations of community variables with intelligence index surpass this value. Highest correlations are with measures of education of the adult population (.36, −.43 and .33), measures of home ownership and its reverse (.37 and −.39) and with indices of the quality and cost of housing (−.30, −.30, and .32). The signs of the correlations depend upon whether the desirable or undesirable end of the dimension being studied corresponds to a large numerical score. Other significant correlations are with proportion of native-born whites (.28), rate of female employment (−.26) and frequency of professional workers (.28). Some of these variables appear to have fairly obvious direct significance, others to have indirect significance as symptoms of some more basic underlying variable.

The prediction obtainable from combinations of variables was studied by determining the multiple correlation for the most promising combinations of two, three, four, five and six variables, following essentially the Wherry-Doolittle procedure. Results are shown in Table III. The multiple correlation increased fairly rapidly, with the addition of variables, until it reached

TABLE I.—INTERCORRELATIONS OF 24 PREDICTOR VARIABLES ($N = 154$)

Variable Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1		21	10	27	−75	23	−06	−48	06	−65	−63	−20	52	−54	−27	−08	−09	−16	−07	12	09	02	−05	−32
2	21		−80	01	−03	07	06	−25	−14	−09	−01	11	28	−24	46	37	41	−16	−33	−03	−17	−25	−31	−29
3	10	−80		16	−22	10	05	16	20	02	−05	−09	−09	06	−46	−34	−38	10	32	22	27	32	32	20
4	27	01	16		−60	93	25	03	70	−05	01	−30	23	−27	−08	−12	−30	09	37	34	27	24	18	−12
5	−75	−03	−22	−60		−53	−23	34	−30	48	47	26	−55	58	47	17	38	−04	−25	−19	−24	−23	−18	22
6	23	07	10	93	−53		24	08	75	03	12	−30	18	−21	01	−05	−19	03	29	37	27	14	10	−10
7	−06	06	05	25	−23	24		29	17	10	16	18	05	−04	−25	−13	−33	39	42	12	20	26	34	10
8	−48	−25	16	03	34	08	29		29	45	62	−01	−62	61	21	02	06	16	13	28	15	01	10	32
9	06	−14	20	70	−30	75	17	29		14	22	−28	−01	−01	−06	−07	−25	10	45	29	28	06	11	−04
10	−65	−09	02	−05	48	03	10	45	14		86	12	−23	25	40	21	28	−05	03	04	05	−33	−18	00
11	−63	−01	−05	01	47	12	16	62	22	86		13	−30	33	45	26	38	−13	−06	12	04	−34	−28	03
12	−20	11	−09	−30	26	−30	18	−01	−28	12	13		−01	00	04	18	19	−12	−15	01	05	−05	04	−06
13	52	28	−09	23	−55	18	05	−62	−01	−23	−30	−01		−98	−31	−02	−06	−19	−01	−14	−10	−10	−10	−49
14	−54	−24	06	−27	58	−21	−04	61	−01	25	33	00	−98		33	06	10	18	−01	12	07	05	06	47
15	−27	46	−46	−08	47	01	−25	21	−06	40	45	04	−31	33		45	68	−23	−45	11	−21	−40	−39	00
16	−08	37	−34	−12	17	−05	−13	02	−07	21	26	18	−02	06	45		44	−25	−34	04	−04	−34	−32	−15
17	−09	41	−38	−30	38	−19	−33	06	−25	28	38	19	−06	10	68	44		−61	−75	08	−23	−57	−63	−21
18	−16	−16	10	09	−04	03	39	16	10	−05	−13	−12	−19	18	−23	−25	−61		66	−09	06	49	57	34
19	−07	−33	32	37	−25	29	42	13	45	03	−06	−15	−01	−01	−45	−34	−75	66		−02	33	54	66	14
20	12	−03	22	34	−19	37	12	28	29	04	12	01	−14	12	11	04	08	−09	−02		62	20	18	−02
21	09	−17	27	27	−24	27	20	15	28	05	04	05	−10	07	−21	−04	−23	06	33	62		27	42	−08
22	02	−25	32	24	−23	14	26	01	06	−33	−34	−05	−10	05	−40	−34	−57	49	54	20	27		70	32
23	−05	−31	32	18	−18	10	34	10	11	−18	−28	04	−10	06	−39	−32	−63	57	66	18	42	70		29
24	−32	−29	20	−12	22	−10	10	32	−04	00	03	−06	−49	47	00	−15	−21	34	14	−02	−08	32	29	

TABLE II.—CORRELATIONS OF PREDICTOR VARIABLES WITH COMMUNITY TEST AND AGE INDICES
(Including Means and Standard Deviations)

Variable	Mean	SD	Correlation			
			IQ	Age Index	Achievement Index	Corrected Achievement Index
1. Population in 1000's	28.9	47.3	−.13	−.14	−.06	−.04
2. Per cent native-born white	86.5	10.6	.28	−.15	−.03	.00
3. Per cent under 5 years old	7.2	1.1	−.04	.29	.07	.01
4. Per cent over 21 years old	66.7	3.4	.19	−.35	−.02	.05
5. Median adult school grade reached	9.3	1.1	.36	−.18	.17	.21
6. Per cent of adults illiterate	8.7	5.8	−.43	.37	.01	−.07
7. Per cent adults graduated from high school	31.2	10.0	.33	−.15	.17	.20
8. Per cent employed males 14 years and over	64.8	6.8	.19	−.13	−.02	.01
9. Per cent employed females 14 years and over	24.7	5.8	−.26	.07	−.03	−.04
10. Professional workers per 1000 population	30.5	13.8	.28	−.18	.21	.25
11. Male domestic servants per 1000 population	1.1	1.9	−.13	.13	.07	.04
12. Female domestic servants per 1000 population	17.6	13.0	−.17	.20	.06	.02
13. Male laborers per 1000 population	23.4	12.3	−.06	.09	−.10	−.12
14. Per cent homes owner-occupied	45.1	10.3	.37	−.18	−.03	.00
15. Per cent homes tenant-occupied	51.2	10.3	−.39	.18	.01	−.03
16. Per cent homes overcrowded	5.6	4.8	−.30	.41	.05	−.03
17. Per cent homes needing major repair	14.4	9.7	−.18	.21	−.13	−.18
18. Per cent homes lacking private bath	29.8	16.4	−.13	.46	−.07	−.18
19. Per cent owner-occupied homes mortgaged	44.2	14.5	.16	−.32	.00	.06
20. Median home rental value (dollars)	25.8	9.5	.32	−.43	.09	.18
21. Per capita retail sales	509.4	190.3	.03	.00	.02	.02
22. Per capita food store sales	122.0	33.2	.07	−.10	.11	.14
23. Per capita county retail sales	326.8	97.8	.14	−.24	.02	.07
24. Average annual pay in retail stores	814.2	161.8	.16	−.30	.04	.11
25. IQ index	100.1	4.0		−.51	.46	.67
26. Age index	49.2	7.3	−.51		.19	
27. Achievement	49.8	4.5	.46	.19		

about .55. Beyond that point the increase was very slow, and the ceiling appears to lie between .55 and .60 for the set of variables used in this study.

Though the correlation found between IQ and community variables is appreciable, it is quite a bit lower than values reported by E. L. Thorndike, Maller, and Davenport and Rem-

TABLE III.—MULTIPLE CORRELATIONS AND REGRESSION WEIGHTS — COMMUNITY FACTORS VS. IQ

Variable	Number of variables analysed				
	2	3	4	5	6
1. Per cent of adults illiterate	−.37	−.21	−.16	−.15	−.13
2. Median home rental	+.23	+.27	+.22	+.22	+.18
3. Per cent homes tenant occupied		−.27	−.30	−.20	−.18
4. Professional workers per 1000			+.13	+.18	+.20
5. Per cent employed females				−.17	−.21
6. Per cent employed males					+.10
Multiple R	.482	.528	.551	.554	.561

mers. There are a number of factors which may have had a part in bringing this about. Mention may be made of the following:

1) Substantial population mobility, due to World War II, between the date of the census (1940) and the date of testing (1945).

2) Non-correspondence between census enumeration units and educational units.

3) Attendance of some children, varying in number and type from one community to another, at private and parochial schools.

4) Failure to include Negro children in the testing program in many of the communities having segregated school systems.

5) Use of small communities with possible lowered reliability of all measures.

These seem all to be factors which would tend to obscure the true relationship between community variables and the intelligence measure. The present results should probably be considered an underestimate of that relationship.

To provide some basis for judging whether the predictors which were most effective in the sample of one hundred fifty-four towns and cities would maintain their effectiveness in another sample of communities they were tested out on twenty-eight county units for which test data were also available. From one point of view the county units are rather unsatisfactory for the 'cross-validation' test, since their characteristics are quite different from those of the urban communities. However, insofar as the same factors are found to be predictive for the counties, the significance of those factors is shown to be more general than would be the case by a test merely in a new group of urban communities.

The most valid single variable, and the most valid combinations of two, three, and four variables were applied to the twenty-eight counties, using the weights shown in Table III. Predicted IQ's were calculated for each county, and these were correlated with the calculated mean IQ. Table IV shows the resulting correlations as well as the mean and standard deviation of each variable.

TABLE IV.—MEAN, STANDARD DEVIATION, AND CORRELATION WITH MEASURED IQ FOR PREDICTED IQ BASED ON COMMUNITY VARIABLES

(Data from 28 County Units)

	Mean	SD	Correlation
Single Most Valid Variable	97.6	3.23	.61
Composite from Two Variables	97.1	2.93	.67
Composite from Three Variables	98.3	3.40	.57
Composite from Four Variables	98.1	3.11	.58
Measured IQ	94.1	6.61	

The results from this analysis are somewhat difficult to interpret, because the county units are substantially more variable in measured IQ than was the case for the urban communities. While the validity of each composite is higher than was the case for the urban communities, the standard error of estimate based on the counties is higher than the standard deviation for the urban communities. Thus, the relative effectiveness of the predictors in the two cases is difficult to estimate. Two things do seem clear. One is that the relative weight to be given to the four variables would be quite different for the county units. This is seen from the fact that the validities resulting from the three and four variable composites are lower than those resulting from only one or two variables. The other is that the rural-urban factor itself, or something intimately associated with it, is a significant factor in determining intelligence test score, over and above the other predictor factors. This is indicated by the difference between mean predicted IQ and mean measured IQ. The measured IQ's averaged three to four points below the predictions. Apparently the ruralness of the county units constituted a handicap of this size.

TABLE V.—MULTIPLE CORRELATIONS AND REGRESSION WEIGHTS COMMUNITY FACTORS VS. ACHIEVEMENT

Variable	Number of Variables Analysed		
	2	3	4
Professional workers per 1000	+.24	+.27	+.26
Per cent homes needing repair	−.16	−.15	−.18
Per cent employed females		−.11	−.16
Male domestic servants per 1000			+.12
Multiple R	.293	.312	.328

Attention may now be turned to the achievement index. Turning again to Table II, it can be seen that the correlations with achievement index are of quite a different magnitude from those with IQ. The highest correlation obtained, after correction for the age factor, was one of .25 with the number of professional workers per 1000 of population. Correlation with median school grade reached by the adult population is .21. These are the only correlations above .20 and significant at the one per cent level. One other measure, per cent of adults who are high-school graduates, yields a correlation of just .20. There is, then, very little evidence that these community variables have any relationship with the achievement index for the community, as measured by the Metropolitan Achievement Test. Multiple correlations with achievement index are shown in Table V. It appears that the ceiling of multiple correlation lies not far above .30.

At first thought, the relatively very low relationships between community factors and the measure of educational achievement seem quite surprising. One argues that if the educational, sociological and economic factors in a community make a difference, this difference should appear more in the frankly learned school achievements than in the general measure represented by an intelligence test. However, several hypotheses may be offered to account for the low relationships.

In the first place, the Metropolitan Achievement Test focuses upon the more conventional and standard objectives of education. It may be that some of the skills which it covers receive substantially less emphasis in the more favored communities, and that for this reason those communities do no better on these skills than the poorer communities. In order to test this, separate analyses were made of two of the Metropolitan subtest scores. One subtest, paragraph reading, was chosen as representing one that it was felt would be heavily emphasized in all schools. Another, arithmetical computation, was chosen because it was considered representative of the type of skill which would be less emphasized in the favored communities. Correlations were computed between these subtest scores and eleven selected predictor variables which had correlated appreciably with IQ. These correlations are shown in Table VI, together with those for IQ.

The results in Table VI are in general accord with our hypothesis. The correlations for paragraph reading correspond roughly to those for IQ. All but one have the same sign, and they show some correspondence in relative size. If the test weights which gave a multiple correlation of .55 for IQ are applied in the case of paragraph reading, the composite correlates .29 with the reading score. In the case of arithmetical computation, however, the pattern of correlations shows no correspondence with that for IQ. In six of eleven cases correlations are actually reversed in sign; that is, factors associated with high community IQ are associated with low community arithmetic scores. The weights which gave a multiple correlation of .55 with IQ yield a correlation of .02 with arithmetic. These results may be explained in terms of a de-emphasis on such skills as arithmetical computation and spelling in the schools of the more advanced small communities. It appears that shifts in educational emphasis have produced a situation in which the more conventional tests of achievement no longer provide an adequate basis for comparative appraisal of communities as a whole, though they may still be useful for studying the individual in relation to his own community. Other types of appraisal are apparently needed if the accomplishments of the schools in outstanding communities are to be assessed adequately.

TABLE VI.—CORRELATIONS OF PREDICTORS WITH IQ, PARAGRAPH READING AND ARITHMETIC COMPUTATION SCORES

Predictor	IQ	Paragraph Reading	Arithmetic Computation
Per cent native-born white	.28	.08	−.16
Per cent over 21	.19	.05	.08
Median school grade reached	.36	.33	.07
Per cent of adults illiterate	−.43	−.18	.09
Per cent employed males	.19	.11	−.04
Per cent employed females	−.26	−.06	.05
Professional workers per 1000	.28	.27	.16
Male domestic servants per 1000	−.13	.02	.06
Per cent homes tenant-occupied	−.39	−.13	.13
Per cent homes needing repair	−.18	−.17	−.13
Per cent homes lacking bath	−.30	−.18	−.17

A second possibility is that the community variables analysed in this investigation are not the ones which are the most important determiners of educational achievement. It may be that if variables dealing with school expenditures, school salaries, library facilities and the like had been included, substantially better prediction of the achievement index would have resulted. Unavailability of this type of information for the many small communities made it impossible to test this hypothesis.

Thirdly, the hypothesis may be advanced that fundamentally education is pretty highly standardized in these United States, and that it is primarily in the non-school factors that communities differ. By this hypothesis, the schools would be thought of as a standardizing influence, and educational achievement would be expected to be less responsive to community variables than would non-school indices. This hypothesis does not lend itself readily to experimental test.

REFERENCES

1) K. S. Davenport and H. H. Remmers. "Factors in state characteristics related to average A-12, V-12 test scores." *J. Educ. Psychol.*, 41, 1950, 110–15.

2) J. B. Maller. "Economic and social correlatives of school progress in New York City." *Teachers College Record*, 34, 1933, 655–70.

3) E. L. Thorndike. *Your City*. New York: Harcourt Brace, 1939.

4) E. L. Thorndike and E. Woodyard. "Differences within and between communities in the intelligence of children." *J. Educ. Psychol.*, 33, 1942, 641–56.

HOW ELEMENTARY-SCHOOL TEACHERS MEET SELECTED SCHOOL SITUATIONS

FRANK SLOBETZ

State Teachers College

St. Cloud, Minnesota

This paper is a résumé of a study made by the writer at the University of Missouri during the school year, 1948–1949. The central problem was to determine what elementary-school teachers did when faced with certain school situations. Included, also, in the study was an attempt to make comparisons of what the elementary-school teachers did on the basis of such teacher characteristics as age, sex, grade levels taught, training, experience, marital status, and parental status. Furthermore, an analysis was made of the teachers' attitudes toward the selected school situations used in the investigation. The chief purpose of the study, therefore, was to determine how competent elementary-school teachers meet selected school situations. More specifically, an attempt was made to answer the following questions:

1) How do the more competent elementary-school teachers of Missouri meet selected school situations?

2) How is the way in which the situations were met related to age, sex, grade level taught, semester hours of training, teaching experience, marital status, and family status?

3) What type of situations do teachers rate the most serious?

4) How are the ratings of the most serious situations related to the several teacher-characteristics classifications?

5) What type of situations do teachers rate the most annoying?

6) How are the ratings of the most annoying situations related to the several teacher-characteristics classifications?

METHOD OF INVESTIGATION

Data were secured by means of an information blank designed to elicit written statements by elementary-school teachers judged by their administrators as the 'more competent' teachers. Seventy per cent of the two hundred eight Missouri public schools of high school district status invited to participate accepted, and nearly seventy per cent of the four hundred twenty-nine blanks sent out were completed and returned. The participating schools from which returns were received were well distributed over the State of Missouri, representing eighty-six of the one hundred fourteen counties in the State.

The elementary-school teachers selected by their administrators as the more competent may be described briefly as follows: Nearly three-fourths of them were thirty-six years old or older. Practically all of them were female teachers. Grouped by grade levels taught, the participating teachers were almost equally distributed in the two levels, grades I–II–III and IV–V–VI, with about one-sixth comprising the grades VII–VIII level. Nearly two-thirds of them fell into the 120–149 semester hours of training group, the remaining one-third being rather uniformly divided among the three other training groups: 60–89, 90–119, 150 and more semester hours. Over three-fourths had more than ten years of teaching experience. Over ninety-five per cent had six or more years of teaching experience. Over one-half of the teachers were married and, of these, 52.1 per cent were parents.

Teachers were requested to write in the space allotted below each of the forty-five situations what they had done about it if they had met it or one similar to it. They were instructed, also, to write 'Have not met' if they had not encountered a situation or one similar to it in their teaching experience.

The forty-five situations were created to include the basic behavior problems reported in Wickman's[1] study. In addition, the list included five exemplary behavioral situations.

The situations, classified by types, are presented below:

Type I. Related to Classroom Decorum.—Fred would not sit still. He was in and out of his seat several times in about as many minutes. Richard played a practical joke on one of the girls, causing general laughter. Mary and Jane were giggling and interfering with the work of the group. You had to leave your room for a few minutes; when you came back you found the children quite noisy and milling about the room. Mary whispered at every opportunity. John and Bill were shooting paper wads (or throwing things) while you were busy with another group. Ann was inclined to monopolize class time by talking about her vacation trip.

---

[1] E. K. Wickman, *Children's Behavior and Teachers' Attitudes*, New York: The Commonwealth Fund, Division of Publications, 1928.

Type II. Related to Authorities or School Regulations.—Johnny was excused to go to the rest room and was gone for twenty minutes. Mary was chewing gum. Jimmy had the habit of throwing his coat on the rack or on the floor. Four of your boys were playing marbles for 'keeps.' Mary was rude to you while you were talking to her about her carelessness. Sam ignored your suggestion that he clean up his desk.

Type III. Related to Application to School Work.—Albert was physically lazy. He showed little interest in his school work. Bill came to school with dirty face and hands. Jack was fifteen minutes tardy one morning. Eddie was careless and untidy in his work. After you explained clearly what the group was to do, Bill later asked you for the same information. George was a non-conformist—he wanted to do things his way.

Type IV. Related to Aggressive and Antagonistic Behavior.—Ann was very critical of others. June dominated her play group, and bossed her classmates with a dictatorial hand. Frank flew into a temper tantrum when he couldn't have his way. Max was a bully; he liked to torment smaller children. Mary liked to tattle on others. John and Jim had a quarrel over sharing some materials. Howard and Sam had a fight on the playground. John deliberately broke his paint brush, issued to him by the school.

Type V. Related to Morality and Integrity.—Jim told highly imaginative tales about his activities. James played with his genital organs during the reading period. You discovered that Mary was cheating on the examination. Johnny took an apple from one of the lunch pails. Betty took a nickel that did not belong to her. Richard used profane language while on the playground. Jack wrote a vulgar note to one of the girls.

Type VI. Related to Withdrawing and Recessive Behavior.—Bob was a shy newcomer. He found it difficult to adjust. Ben was a physical coward; he was afraid to stand up for himself. Mark was a quiet boy and never bothered anyone. Eleanor was suspicious of her classmates. She told you that they were talking about her. Richard's feelings were hurt because he was not chosen to serve as librarian.

Type VII. Related to Exemplary Behavior.—Sam held the door open for his classmates as they came in the building. A group of your boys volunteered to lend their basketball to the girls. John found a dollar bill in the school building and brought it to you. Mary told you that it was her fault that the scissors were lost and that she was sorry. Jane shared her new story book with her classmates.

Since teachers were asked to report actual treatment of the situations in an open-question type of inquiry, it was necessary to analyze the statements to discover the basic techniques involved. The pattern of treatments presented in this study evolved outrightly from this method of treating the data provided by the teachers.

FINDINGS

In the original dissertation, the writer organized the findings by devoting one chapter to each of the seven types of situations. Each chapter presented one comprehensive table showing specifically how all teachers met each of the indicated situations. Included, also, in each chapter were seven other tables which showed comparative percentages by teacher-characteristics classification of how the elementary-school teachers met the situations related to the indicated types. One chapter was devoted to the attitudes of teachers toward the forty-five situations.

For purposes of this paper, some of the data have been re-organized to present in briefer form certain of the more important findings. Table I shows how two hundred ninety elementary-school teachers met the forty behavior-problem situations. One hundred thirteen specific measures, classified under thirteen types, were reported employed by the teachers. On a punitive-constructive dichotomy basis, Table I reveals that slightly more than one-fourth of all measures used were punitive in nature, and slightly less than three-fourths were constructive in nature. Physical Force, Censure, Overtime or Extra Work, Deprivation, Sent or Referred to Office, and Penalties were considered punitive in nature. Ignored or Did Nothing, Verbal Appeal, Group Reaction, Constructive Assistance, and Commendation were considered constructive in nature. Rectification or Reparation was considered neither.

TABLE I.—HOW THE 290 ELEMENTARY-SCHOOL TEACHERS MET THE FORTY BEHAVIOR-PROBLEM SITUATIONS

(Figures in parentheses give frequency)

How Situations Were Met	Per cent
Physical Force (134): Spanked (106), Shook (15), Tied in seat (6), Slapped (5), Obstructed pupil's desk (1), Held pupil (1)	1
Censure (1088): Scolded (173), Commanded (114), Warned (68), Refused to listen (62), Used saturation (56), Refused to help (54), Secured apology (52), Reciprocated (45), Shamed (36), Questioned (36), Looked at (34), Identified publicly (29), Indicated disappointment (26), Required sharing with group (23), Destroyed item (21), Refused to accept work (20), Embarrassed (17), Used sarcasm (17), Required pupil to stand near his seat (15), Lectured (14), Required repeat writing (14), Stood near pupil (13), Refused to recognize (13), Indicated privately discovery (10), Shook head in disapproval (10), Seated pupil near teacher (9), Required heads on desks (8), Threatened (8), Demanded respect (8), Required standing in corner (8), Stopped and waited (8), Soaped mouth (7), Ridiculed (6), Sent home (6), Hushed (5), Secured confession (4), Displayed poor work (4), Scoffed at (4), Indicated displeasure (4), Kept list of names (4), Made comparison (4), Laughed at (4), Required repeating to teacher (3), Impersonated pupil (3), Indicated surprise (2), Sent for parent (2), Required keeping coat on (1), Required use of broken item (1), Required covering of eyes with hands (1), Required putting glue and cotton on fingers (1), Required pupil to stick gum on end of nose (1).	10
Overtime or Extra Work (486): Required to repeat work (189), Required making up time lost (121), Kept in (91), Assigned extra work (70), Required to clean up (15).	4
Deprivation (1044): Deprived recreational time (223), Isolated (171), Rearranged seating (150), Denied privilege (112), Denied use of materials (95), Omitted from group activity (94), Assigned less attractive work (74), Restricted participation (68), Removed from class (41), Took away article or item (16).	10

TABLE I.—*Continued*

How Situations Were Met	Per cent
Sent or Referred to Office (65):	0.6
Penalties (99):	1
Demerits (95), Money fines (3), Non-promotion (1).	
Rectification or Reparation (214):	2
Required replacement (150), Required payment (56), Required giving up of personal article (6), Required double replacement (1), Paid for by teacher (1).	
Ignored or Did Nothing (512):	5
Verbal Appeal (2586):	26
Used reasoning (1771), Reminded (318), Made simple inquiry (218), Requested cessation (162), In behalf of the group (109), In behalf of teacher's regard (8).	
Group Reaction (143):	2
Constructive Assistance (3167):	31
Tried to create opportunity for successful participation (638), Tried to provide appropriate activity (338), Helped through pupil assistance (279), Suggested Correction (207), Conferred with or contacted parent (206), Created situation to provide insight (185), Varied procedure or participation (152), Talked to group about situation or condition in general (151), Tried to provide incentives (148), Accepted situation with group (132), Served as arbiter (96), Assigned special responsibility (95), Respected his ideas and held him responsible for group coöperation (82), Repeated directions (81), Helped directly to meet it (63), Tried to arouse emulation (57), Arranged for play with his peers (50), Suggested hand activity (43), 'Kidded' pupil, (33), Repeated later (24), Listened to pupil (17), Saw that bully met his match (16), Instituted immediate class work (10), Suggested pupils work it out amicably (8), Organized a club (8), Held frank discussion (7), Assigned committee work (7), Secured reconciliation (6), Suggested another time and place (6), Created attention-getting situation (8), Deferred action (3), Suggested substitute (3), Provided protection for pupil (3).	
Commendation (57):	0.5
Personal (39), Public (11), Used as teaching situation (4), Via report card (3).	

TABLE I.—*Continued*

How Situations Were Met		Per cent
Searched for Reasons of Behavior (620)		6
Tried Many Things Unsuccessfully (29)		0.3
Total Number of reports on situations met	(10244)	99.4
Reports of non-encounter	(1356)	
Total number of reports	(11600)	

Table II presents a comparison of how the teachers—classified by age, sex, grade levels taught, semester hours of training, years of teaching experience, marital status, and parental status—met the forty behavior-problem situations. The percentages shown were computed on the basis of the number of situations encountered.

Percentage differences within each classification and for each type of situation and for all situations considered together were treated statistically for significance. Critical ratios were determined by dividing the differences by the respective standard errors of the differences. No critical ratios indicating significance at levels ranging from five per cent to one per cent were found for any of the differences between percentages shown in Table II. Critical ratios approaching significance were found as indicated below:

Type I	
(25 or Less)–(26–35)	CR: 1.4
(46 or More)–(26–35)	CR: 1.6
(120–149)–(90–119)	CR: 1.4
Type V	
(1-2-3)–(7–8)	CR: 1.4

On the basis of Tables I and II, the following summary statements may be made:

1) In meeting situations related to Classroom Decorum, teachers used constructive measures slightly more frequently than they employed punitive ones.

2) In meeting situations related to Authorities or School Regulations, teachers employed constructive methods about twice as frequently as they used punitive measures.

TABLE II.—A COMPARISON OF HOW 290 ELEMENTARY-SCHOOL TEACHERS MET FORTY BEHAVIOR-PROBLEM SITUATIONS IN TERMS OF PERCENTAGES OF CONSTRUCTIVE MEASURES USED

Classification of Teachers	No. of Teachers in Each Class	Per cent of Constructive Measures Used According to the Six Types of Situations						
		I	II	III	IV	V	VI	All Types
Ages								
25 or Less	11	64	61	80	67	71	97	73
26–35	70	48	67	79	61	62	96	69
36–45	120	53	65	80	59	67	96	70
46 Plus	89	56	66	77	62	68	96	71
Sex								
Female	277	53	65	79	60	68	97	70
Male	13	60	68	72	72	64	95	72
Grade levels taught								
1-2-3	117	54	69	83	60	71	96	72
4-5-6	127	52	64	77	62	66	96	70
7-8	46	52	60	74	61	59	96	67
Semester hours of training								
60–89	34	49	64	80	58	63	96	68
90–119	38	47	60	74	64	66	97	68
120–149	181	59	66	79	60	67	97	71
150 Plus	37	54	69	80	66	72	98	73
Years of teaching experience								
2–5	12	67	63	78	56	65	97	71
6–10	52	52	64	76	61	71	96	70
10 Plus	226	53	66	80	61	66	96	70
Marital status								
Single	123	54	67	80	62	60	96	72
Married	167	53	65	77	60	65	96	69

TABLE II.—*Continued*

Classification of Teachers	No. of Teachers in Each Class	Per cent of Constructive Measures Used According to the Six Types of Situations						
		I	II	III	IV	V	VI	All Types
Married								
With Children	87	50	61	77	59	62	97	68
Without Children	80	55	67	79	59	68	97	71
All Teachers		54	66	79	61	67	97	71

Type I Seven situations related to Classroom Decorum.
II Six situations related to Authorities or School Regulations.
III Six situations related to Application to School-work.
IV Eight situations related to Aggressive and Antagonistic Behavior.
V Seven situations related to Morality and Integrity.
VI Six situations related to Withdrawing and Recessive Behavior.

3) In meeting situations related to Application to School Work, teachers used constructive measures about four times as often as they used punitive measures.

4) In meeting situations related to Aggressive and Antagonistic Behavior, teachers used constructive measures about twice as often as punitive ones.

5) The constructive methods employed by all teachers in meeting situations related to Morality and Integrity constituted nearly seventy per cent of all measures used.

6) Nearly ninety-seven per cent of the measures used by teachers in meeting situations related to Withdrawing and Recessive Behavior were constructive in nature.

7) In terms of teacher-characteristics classifications, no significant differences were found in how teachers met the situations, considered by types or by all situations as a whole.

8) The two hundred ninety teachers made 11,600 reports on how they met the forty behavior-problem situations. Of this number, eleven per cent were of non-encounter. Types of measures employed by all teachers in connection with the forty situations, in order of frequency of use, were: Constructive Assistance (thirty-one per cent), Verbal Appeal (twenty-six per cent), Censure (ten per cent), Deprivation (ten per cent), Searched for Reasons of Behavior (six per cent), Ignored or Did Nothing (five per cent), Overtime or Extra Work (four per cent), Rectification or Reparation (two per cent), Group Reaction (two per cent), Physical Force (one per cent), Penalties (one per cent), Commendation (0.5 per cent), and Sent or Referred to Office (0.6 per cent). Slightly more than one-fourth of the measures employed in meeting the forty situations were punitive in nature and slightly less than three-fourths were constructive in nature.

In order to assemble pertinent data about the treatment of the situations related to Exemplary Behavior, Table III was constructed. How the teachers handled these situations is reported in terms of six categories and in terms of frequency by per cents. As Table III indicates, teachers used the Commendation technique ninety-four per cent of the time. In terms of teacher-characteristics classifications, no significant differences were found in the use of the Commendation technique. A significant difference (CR: 2.64) at one per cent level was found for Ignoring or Doing Nothing between teachers with ten years or more teaching experience and those with two to five years of experience.

ATTITUDES

In addition to the written reports on how they had met the forty-five situations, the participating teachers were asked to indicate the most serious, the second most serious, and the third most serious situations. They were also asked to indicate the most annoying, the second most annoying, and the third most annoying situations. In effect, the teachers were asked to indicate rank orders from one through three for the three most serious and the three most annoying situations. Two hundred eighty teachers indicated their attitudes in the manner prescribed. In order to consolidate the three attitudinal expressions

TABLE III.—A COMPARISON OF HOW 290 ELEMENTARY-SCHOOL TEACHERS MET FIVE SITUATIONS RELATED TO EXEMPLARY BEHAVIOR

Classification of Teachers	No. of Teachers in Each Class	How Teachers Met Situations and Per Cents of Frequency					
		A	B	C	D	E	F
Ages							
25 or Less	11	0	12	2	0	0	86
26–35	70	0.3	4	1	0.6	1	93
36–45	120	1	1	1	0.2	0.5	96
46 Plus	89	1	3	1	0.2	0.2	94
Sex							
Female	277	1	2	1	0.3	0.6	95
Male	13	0	3	0	0	0	97
Grade Levels Taught							
1-2-3	117	0.7	3	0	0	0.4	96
4-5-6	127	1	4	2	0.7	0.5	92
7-8	46	0.5	3	2	0	1	93
Semester Hours of Training							
60–89	34	0	4	3	0.6	0.6	92
90–119	38	0	4	0	0.6	0	95
120–149	181	1	3	1	0.2	0.7	94
150 Plus	37	0.7	2	1	0	0.7	95
Years of Teaching Experience							
2–5	12	0	0	2	0	0	98
6–10	52	0	6	1	0.4	0.4	92
10 Plus	226	1	3	1	0.3	0.7	94
Marital Status							
Single	123	0.9	2	2	0.5	0.7	94
Married	167	0.9	4	1	0.1	0.5	94
Married							
With Children	87	0.8	3	1	0.3	0.3	94
Without Children	80	1	4	0.8	0	0.8	93
All Teachers		0.9	2.7	1.	0.3	0.6	94

Key: A—Rectification or Reparation. B—Ignored or Did Nothing. C—Verbal Appeal. D—Group Reaction. E—Constructive Assistance. F—Commendation.

into a single index, the three ranks were weighted by multiplying the frequencies for ranks 1, 2, and 3 by 3, 2, and 1, respectively.

These attitudinal indices were converted into percentages as Table IV indicates. This table reveals that slightly over one-half of the ratings as to seriousness were directed at situations related to Morality and Integrity. About one-third of the ratings were divided almost equally between situations related to Aggressive and Antagonistic Behavior and situations related to Withdrawing and Recessive Behavior.

Table IV also reveals that over one-third of the ratings as to annoyance involved situations related to Classroom Decorum. Situations related to Application to School Work and those related to Aggressive and Antagonistic Behavior were rated almost as annoying as the Classroom Decorum type.

TABLE IV.—PERCENTAGE DISTRIBUTION OF 280 ELEMENTARY-SCHOOL TEACHERS' RATINGS OF THE THREE MOST SERIOUS AND THE THREE MOST ANNOYING SITUATIONS CLASSIFIED BY TYPES OF SITUATIONS

Types of Situations	Most Serious	Most Annoying	CR
Situations Related to Morality and Integrity	50.4	3.0	14.8
Situations Related to Aggressive and Antagonistic Behavior	17.1	24.9	2.27
Situations Related to Withdrawing and Recessive Behavior	16.1	5.6	4.0
Situations Related to Application to School Work	13.1	26.5	4.0
Situations Related to Authorities or School Regulations	2.1	4.3	1.4
Situations Related to Classroom Decorum	1.2	35.7	17.0
Total Per Cents	100.0	100.0	

The critical ratios shown in the table were computed by dividing the percentage differences between corresponding types of situations by the respective standard errors of the differences. All are significant at the one per cent level, except two; and one of these approaches significance at the two per cent level. These statistics indicate that the teachers considered situations related to Morality and Integrity and to Withdrawing and Recessive Behavior more serious than annoying; and situations related to Classroom Decorum, to Application to School Work, and to Aggressive and Antagonistic Behavior as more annoying than serious.

For comparative purposes, Table V was constructed. Teachers' ratings of the three most serious situations were converted to percentages and compiled by classifications of teachers and by types of situations. Statistical treatment for significance of percentage differences within each classification and for each type of situations produced the following significant critical ratios:

Type V	CR	Level
(60–89)—(150 or more)	1.95	5%
(90–119)—(150 or more)	2.5	2%
(120–149)—(150 or more)	2.1	5%
Type VI		
(26–35)—(25 or less)	2.34	5%
(36–45)—(25 or less)	2.68	1%
(46 or more—(25 or less)	2.40	2%
(150 or more)—(60–89)	2.55	2%
(150 or more)—(90–119)	2.10	5%
(150 or more)—(120–149)	2.38	5%

With respect to situations related to Morality and Integrity (Type V), teachers with semester hours of 60–89, 90–119, and 120–149 considered situations of this type more serious than teachers with 150 or more semester hours.

Critical ratios for situations related to Withdrawing and Recessive Behavior (Type VI) indicate that the older groups of teachers considered this type of situation more serious than the youngest age group, 25 or less. Significant differences were also found for Type VI in the semester hours of training classification. Teachers with 150 or more hours rated these situations more serious than teachers in the three other groups.

As Table VI shows, teachers' ratings of the three most annoying situations were converted into percentages and compiled by

TABLE V.—PERCENTAGE DISTRIBUTION OF 280 ELEMENTARY-SCHOOL TEACHERS' RATINGS OF THE THREE MOST SERIOUS SITUATIONS CLASSIFIED BY TYPES OF SITUATIONS

Classification of Teachers	Number of Teachers	Percentages for Each Type Situation					
		I	II	III	IV	V	VI
Ages:							
25 or Less	10	0	3.3	6.7	20	68.3	1.7
26–35	66	1.3	4.3	17.4	15.4	45	16
36–45	118	.4	1.7	13.1	16.8	51	17
46 Plus	86	2.2	1.4	11.3	17.6	50	16
Sex							
Female	267	1.2	1.8	13.2	17.6	50	16.4
Male	13	0	12.3	19.2	19.2	31.5	17.8
Grade Levels Taught							
1-2-3	111	1.7	1.8	13	17.7	49.5	16.4
4-5-6	125	.7	2.4	12.5	16.6	50.5	17.4
7-8	44	1.2	3.1	16.8	16.9	50	11.9
Semester Hours of Training							
60–89	33	1	2.6	14.2	17.8	54.8	9.6
90–119	39	0	7.8	9.1	10	59.7	13.4
120–149	173	.9	1.8	15.7	17.7	50.2	13.8
150 Plus	35	3.8	2.4	6.7	21.4	31.9	33.8
Years of Teaching Exp.							
2–5	12	0	8.2	17.8	17.8	45.2	11
6–10	52	1.6	4.2	11.2	16	52.2	14.8
10 Plus	216	1	1.5	13.9	17.2	49.7	16.7
Marital Status							
Single	121	.7	2.4	15.7	19.6	44.2	17.5
Married	159	1.6	2.2	11.7	15.2	54.2	15.1
Married							
With Children	83	2.5	1.4	10.4	14.5	56	15.1
Without Children	76	.9	3.1	13.2	15.6	52	15.2

Table VI.—Percentage Distribution of 280 Elementary-school Teachers' Ratings of the Three Most Annoying Situations Classified by Types of Situations

Classification of Teachers	Number of Teachers	Percentages for Each Type Situation					
		I	II	III	IV	V	VI
Ages:							
25 or Less	10	38.3	10	25	21.7	5	0
26–35	66	37.4	6.4	28.2	20.8	1.6	5.6
36–45	118	32.4	3.3	29.2	25	3.4	6.1
46 Plus	86	37	3.4	23.1	27.9	3.4	5.2
Sex:							
Female	267	36.1	4.1	26.6	25.1	2.5	5.6
Male	13	31	7	21.1	23.9	14.1	2.8
Grade Level Taught							
1-2-3	111	34	4.5	24.5	27.2	3.3	6.5
4-5-6	125	36.4	4.1	27.3	23.5	3.5	5.2
7-8	44	41.1	4.4	25.8	23.8	.8	4.1
Semester Hours of Training							
60–89	33	46.6	4.7	23.6	16.2	1.6	7.3
90–119	39	40	5.7	23.2	30.7	3.5	.9
120–149	173	36.1	4.5	28.5	22.3	2.2	6.7
150 Plus	35	31.4	1.5	20	36.3	7.8	2.9
Years of Teaching Experience							
2–5	12	42	7.3	23.2	24.6	2.9	0
6–10	52	42.1	5	22.8	25.5	1	3.6
10 Plus	216	34.1	3.7	25.2	27	3.5	6.3
Marital Status							
Single	121	34.7	5.2	25.3	27.6	1.8	5.4
Married	159	36.4	3.6	27.5	23.1	3.8	5.6
Married							
With Children	83	41.5	3.5	24.6	20.9	3.9	5.6
Without Children	76	31.3	3.8	30.4	25.3	3.6	5.6

classifications of teachers and by types of situations. Statistical treatment for significant percentage differences indicates the following critical ratios:

	CR	Level
Type IV		
(150 or more)—(60–89)	1.94	5%
Type VI		
(36–45)—(25 or less)	2.76	1%
(46 or more)—(25 or less)	2.17	5%
(10 years or more)—(2–5 years)	3.80	1%

Teachers with 150 or more semester hours considered situations related to Aggressive and Antagonistic Behavior more annoying (approaching significance) than teachers with 60–89 semester hours.

The most significant critical ratios were found in connection with situations related to Withdrawing and Recessive Behavior. Teachers in age classification, 36–45, considered these situations more annoying than teachers of 25 or less, as did the oldest group of 46 or more. The most experienced teachers, 10 years or more, considered this type more annoying than teachers with two to five years of experience.

RANK ORDER DATA

Teachers' ratings of the most serious situations in rank order are as follows: (1) stealing nickel, (2) sex irregularity, (3) cheating on examination, (4) vulgar note, (5) daydreaming, (6) temper tantrum, (7) laziness, (8) bully, (9) non-conformity, (10) stealing apple, (11.5) suspiciousness, (11.5) profanity, (13) quiet behavior, (14) shyness, (15) domination, (16) cowardice, (17) critical behavior, (18) inattention, (19) rudeness, (20.5) carelessness, (20.5) marble playing, (22.5) noisiness, (22.5) property destruction, (24) fighting, (25) imaginative tales, (26) tattling, (27) monopolizing class time, (29.5) overactivity, (29.5) chewing gum, (29.5) ignoring teacher, (32) quarreling.

Teachers' rating of the most annoying situations in rank order are as follows: (1) inattention, (2) tattling, (3) whispering, (4) overactivity, (5) giggling, (6) daydreaming, (7) noisiness, (8) laziness, (9) bully, (10) temper tantrum, (11) non-conformity,

(12) paper wads, (13) carelessness, (14) domination, (15.5) chewing gum, (15.5) critical behavior, (17) rudeness, (18) monopolizing class time, (19) cheating on examination, (20) suspiciousness, (21) tardiness, (22) fighting, (23) property destruction, (24) profanity, (25) vulgar note, (26.5) quarreling, (26.5) throwing coat on floor, (29) ignoring teacher, (29) sex irregularity, (29) practical joke, (31) undue absence from class, (32) cowardice, (34.5) uncleanliness, (34.5) imaginative tales, (34.5) hurt feelings, (34.5) stealing nickel.

The rank-difference correlation between the two rank-order ratings is .02.

CONCLUSIONS

On the basis of the findings and within the framework of limitations indicated in the study, the more important conclusions seem to be those presented below.

1) In meeting behavior-problem situations, the designated competent elementary-school teachers operated from a constructive point of view, with some evidence of searching for the reasons of behavior when certain situations were faced.

2) Situations of the Exemplary type were generally met by the designated competent teachers with specific forms of commendation.

3) Teachers employed the highest percentages of constructive measures in connection with Withdrawing and Recessive Behavior type of situations, and the lowest with situations related to Classroom Decorum.

4) In terms of the seven teacher-characteristics classifications, no significant differences were found in the way teachers treated the situations, either by types of situations or by all forty behavior-problem situations considered as a whole.

5) The teachers were most seriously concerned with situations related to Morality and Integrity. About one-sixth of their ratings included situations related to Aggressive and Antagonistic Behavior and about the same portion of the ratings were directed at situations related to Withdrawing and Recessive Behavior.

6) The three chief types of annoying situations were those classified as Classroom Decorum, Application to School Work, and Aggressive and Antagonistic Behavior.

7) Situations related to Morality and Integrity and to Withdrawing and Recessive Behavior were considered more serious than annoying; and those related to Classroom Decorum, to Application to School Work, and to Aggressive and Antagonistic Behavior as more annoying than serious.

8) Teachers with the greatest amount of training considered situations related to Morality and Integrity less serious than teachers with lesser amounts of training.

9) The oldest group of teachers considered Withdrawing and Recessive Behavior situations more serious than the youngest group. Teachers with the most training considered these situations more serious than teachers with less training.

10) Teachers with most training considered situations related to Aggressive and Antagonistic Behavior as more annoying than teachers with the least training.

11) The two older groups of teachers considered situations related to Withdrawing and Recessive Behavior more annoying than the youngest group of teachers. The most experienced teachers considered these situations more annoying than the least experienced teachers.

12) The findings suggest that elementary-school teachers show more real differences in their attitudes toward behavior of children than in their treatment of such behavior.

THE RELATIONSHIP EXISTING BETWEEN BILINGUALISM AND RACIAL ATTITUDE*

GRANVILLE B. JOHNSON, JR.

Emory University

In many regions of the southwest United States there are groups of people who find it necessary to use two languages for normal social intercourse: Spanish in the home and English in schools and in business transactions. As in most bilingual situations, one language is given more social acclamation than the other, since it is representative of the more dominant culture.

The study of language may be much more than language study, *per se*, since it is justifiable to assume that language is a reflection of the personality structure of the individual.[1] If personality does, as Frank assumes, "organize experience in terms of its own private idiosyncratic world of meanings,"[2] one may conclude with him that it is logical to extend this assumption to include an individual's choice of language for self-expression as one aspect of personality.

The purpose of the present investigation was to determine the relationship existing between the bilingual background of Spanish subjects and the attitude of the Spanish toward the Anglo ethnic group.

SUBJECTS AND MATERIALS

Materials.—

1) The Projective Test of Racial Attitudes composed of six carefully selected pictures relevant to race situations and provocative of projected response was utilized for the analysis of racial attitudes.

2) A tabulation sheet was utilized for recording the dynamics of responses.

3) The Hoffman Bilingual Schedule

---

* This is the final article of a series of three concerned with the origin and development of racial attitude. The first, "An Experimental Projective Technique for the Analysis of Racial Attitude," appeared in April, 1950, issue (41: 4). The second, "The Origin and Development of the Spanish Attitude Toward the Anglo and the Anglo Attitude Toward the Spanish," was published in May, 1950 (41: 5).

Subjects.—

1) Anglo: thirty male subjects (four-, eight- and twelve-year levels, N = 90) from public schools in a southwestern town.

2) Spanish: thirty bilingual male subjects (four-, eight- and twelve-year levels, N = 90) from public schools in a southwestern town.

A further limitation was that members of both groups were required to have lived in this specific town for the year prior to testing.

METHODS OF MEASURING BILINGUALISM

Means of measurement of bilingualism may be classified into three categories: (1) Background questionnaires, (2) Association techniques, and (3) Rating scales.

1) *Background Questionnaires.*—Prescott, in 1932, described a bilingual questionnaire of twenty items which covered questions concerning oral and written language influence in the child's family background.[3] He assigned a numerical value to each significant item; the sum of these values represented the individual's foreign language index. Other questionnaires of similar character (e.g., Hoffman, 1933) have developed since this time.

The technique employed in the present investigation utilized the Hoffman Bilingual Schedule.

2) *Association techniques.*—M. Gali of Spain has proposed two methods for the measurement of degree of bilingualism. The first of these methods is a test of immediate verbal memory. Two parallel series of ten words—ten in Spanish and ten in Catalan—are orally presented to the child, each series being read three times with a fixed time interval in between the words. At the end of the presentation the children are requested to reproduce on a piece of paper the words which they are able to recall.[4]

Gali utilized the chain association method in which he presented eight words in Spanish given alternately with eight in Catalan. The subject responded with any word in either of the two languages which occurred to him within a fixed time.[5]

Saer employed an association method to measure the affective value of words in the two languages of the bilingual individual. Fifty words representing situations, persons, objects and actions that normally are part of a child's world at the age of three are selected. These fifty words in English and their equivalents in Welsh . . . without any definite order of presentation as to the English or Welsh words are given to the child individually. The responses given in either language are noted. Account is taken of the association time for each word in each of the two languages. The association time for the Welsh word is divided by the time of the association for its equivalent in English. This quotient is equal to 100 when the association times for the two responses are equal. A score of less or more than 100 will indicate the greater or smaller affective value of the situation, object, person, or action in one or the other of the two languages. A quotient of approximately 100 will show that the child is truly bilingual in reference to the particular situation, object or person denoted by the word.[6]

3) *Rating Scales.*—Degree of language assimilation represents a continuum extending from zero to one hundred per cent. Malherbe divided this continuum into six progressively comprehensive stages, thus making it possible to compare roughly the two languages of a bilingual individual. In the study the Hoffman Bilingual Schedule was employed to measure the bilingual background of the Spanish subjects.

The test in its final form presented the following questions to which five responses were possible: Never, Sometimes, Often, Mostly, Always:

1) Do the following speak to you any language other than English? (a) Father. (b) Mother. (c) Grandfather. (d) Grandmother. (e) Brothers and Sisters. (f) Relatives.

2) Do you speak to the following any language other than English? (a) Father. (b) Mother. (c) Grandfather. (d) Grandmother. (e) Brothers and Sisters. (f) Relatives.

3) Does your father speak to the following any language other than English? (a) Mother. (b) Brothers and Sisters.

4) Does your Mother speak to the following any language other than English? (a) Father. (b) Brothers and Sisters.

5) Do your Brothers and Sisters speak to the following any language other than English? (a) Father. (b) Mother.

6) Do the following read any newspaper in a language other than English? (a) Father. (b) Mother. (c) You.

7) Do the following read any books in a language other than English? (a) Father. (b) Mother. (c) You.

8) Do the following write any letters in a language other than English? (a) Father. (b) Mother. (c) You.

9) Are letters written in a language other than English received in your home?

10) Do the following attend lectures given in a language other than English? (a) Father. (b) Mother. (c) You.

11) Do the following attend the theatre where plays are given in a language other than English? (a) Father. (b) Mother. (c) You.

12) Are radio programs which are given in a language other than English listened to in your home?

13) Do you do your thinking in a language other than English?

14) Are there any books in a language other than English in your home? [This question is answered: (a) None. (b) Some. (c) Many. (d) Most. or (e) All.]

The major conclusions Hoffman drew from analysis of the test are in summary:

Validity, found by correlating the bilingual scores against ratings as to the extent of bilingual background is .73 = .03

Coefficient of correlation between bilingual scores and estimates of bilingual background based on interviews is .82 = .03

Validity was established by demonstrating that the test differentiates between groups whose extent of bilingual background is relatively unknown.

Retest reliability coefficient (interval of three weeks) is .81 = .02

Reliability (split-halves) is .845. When corrected by Spearman-Brown, the coefficient is .92 for the total schedule.

There appeared to be no relationship between bilingual background and chronological age or with grade status for ages ten to fourteen and grades V to VIII.

No sex difference in regard to either size or variability of the bilingual scores was reported.

This test was administered to each Spanish subject at the time of the presentation of the Projective Test of Racial Attitudes. It was found, early in the investigation, that the Spanish administrator was more satisfactory than the Anglo. This was in part because of his knowledge of the Spanish language and culture.

PROCEDURE

The purpose of the present investigation was to determine the relationship existing between the bilingual background of the selected Spanish subjects and the attitude of the sample Spanish at the chosen age levels toward the Anglo ethnic group.

The method of procedure used to ascertain this relationship was to derive the prejudice score from each of the Spanish protocols of the Projective Test of Racial Attitudes and correlate it with the bilingual score of the Hoffman Bilingual Scale.

RESULTS

Table I presents the mean and standard deviation of the selected Spanish subjects on the Hoffman Bilingual Scale.

TABLE I.—THE MEAN AND STANDARD DEVIATION OF THE SPANISH SUBJECTS ON THE HOFFMAN BILINGUAL SCALE

		M	σ_m	σ
Age Levels	12	181.001	9.078	48.885
	8	157.900	10.382	55.905
	4	104.512	9.501	51.165

Though the technique for establishing bilingual background was identical throughout the age groups, the Spanish at the twelve-year level were revealed to possess a greater (though not significantly so) bilingual background than subjects at the eight-year level, and the eight-year was greater than the four. It is possible to explain this feature on the basis that, though the frame of reference set up for the study excluded all subjects who had not lived in the community for a year, this at the four-year level represented a minimum of one-fourth the total experience of the subject. This weighting decreased progresssively with age. It was more possible, therefore, at the older levels, for the subject to have spent eleven-twelfths of his life in rural districts with their concomitant emphasis on the Spanish language.

The relationship existing between the bilingual score and the total prejudice score yielded the results found in Table II. It is observable from this table that the rectilinear relationship existing between bilingual background and prejudice increases progressively with age. The eta correlations were made to ascertain if any curvilinear functions were present. It may be noted that the *Nxy* did not deviate significantly from rectilinearity, but the *Nyx* demonstrated that, at the four-year level, decreasing progressively through the eight and twelfth years,

TABLE II.—DEGREE OF CORRELATION BETWEEN THE BILINGUAL SCORE AND TOTAL PREJUDICE SCORE OF THE RACIAL ATTITUDES TEST FOR EACH OF THE SPANISH AGE LEVELS
N = 30 FOR EACH AGE LEVEL

Age Level	r	σ_r	Nyx	σ_{Nyx}	Cor. Nyx	x^2 Nyx	Lev. of Sig. x^2Nyx	Nxy	σ_{Nxy}	Cor. Nxy	x^2 Nxy	Lev. of Sig. x^2Nxy
12	.2600	.1731	.4433	.7640	.1492	.0772	−.5180	.1680	8.9660	15.3381	.50	.50
8	.1687	.1804	.6421	.7320	.1091	.0862	.3442	.2676	11.1161	13.1160	.40	.50
4	−.1706	.1820	.7520	.6401	.0807	.1072	.5943	.1158	23.2700	8.1741	.02	.90

there is a tendency for those subjects with least prejudice to have the highest and lowest bilingual scores.

Though this curvilinear function at the four-year level is possibly reflective of parental disposition, it indicates, nevertheless, that a profound knowledge of the Anglo culture or no knowledge of it at all yields least bias toward it. Possibly, with less bilingual background, the Spanish become part of the Anglo culture and in many cases identify themselves with it. This tendency might explain the finding previously discussed for many of the younger Spanish subjects to be prejudiced in favor of the Anglo.

To analyze this problem in greater detail, the degree of correlation between the bilingual score and each of the following categories was computed: Effect of Environment, (Frustrating); Reaction to Environment, (Frustration); Adequacy of Principle Character, (Subordinate); Ending Hero, (Defeat); Ending Hero, (Victory); and Ending Theme, (Unsatisfactory).

These correlations bore out the findings of the total analysis: the rectilinear relationship between bilingual background and prejudice tends to increase with age; but at the four-year level, decreasing progressively through the eighth and twelfth years, those subjects with the highest and lowest bilingual scores have the least prejudice.

At the four-year level, the rectilinear relationship between prejudice and bilingual background was negative, the *Nxy* is insignificant, and the *Nyx* is at the .30 level of significance. This last finding would indicate that those individuals at the four-year level with the most and least bilingual background had the least tendency to end the theme unsatisfactorily. At the eight-year level, however, those subjects with the highest and lowest bilingual backgrounds had equal prejudice. At the twelve-year level the curvilinear (*Nyx*) relationship (significant at the .01 level) was the reverse of the findings at the four-year level which would indicate that, at the oldest age level, those individuals with the greatest and least bilingual backgrounds tend to end the theme unsatisfactorily between the Spanish and Anglos. This finding could possibly indicate that the twelve-year Spanish with low bilingual backgrounds realize the barrier between the Spanish and Anglos and consider resolving it with pessimism. Those Spanish with intensive bilingual backgrounds, however, may end the theme unsatisfactorily at this level as a result of increased social intercourse with the Anglos who are quite prejudiced at this age. This contact touches the Spanish ego, and ending the theme unsatisfactorily may be a negativistic device utilized to resolve tension.

The rectilinear correlation which shows greatest increment from younger to older levels is that concerned with the adequacy of the principle character. This would indicate that, with age, the relationship between bilingual background and feelings of superiority increases. It is possible that feelings of 'national' pride along with ego involvement and identity may explain this finding.

The rectilinear correlations which deviate from the general trend are found only in Adequacy of Principle Character, (Subordinate) and in Ending Hero, (Defeat). This would indicate that as age increases there is less tendency for subordination and defeat of the hero to be associated with a high bilingual background score. This factor, too, may be explained on the basis of pride resulting from frustration.

SUMMARY AND CONCLUSIONS

The purpose of the present investigation was to determine the relationship existing between the bilingual background of the Spanish subjects and the attitude of the sample Spanish at the chosen age levels toward the Anglo ethnic group.

The method of procedure used to ascertain this relationship was to derive the prejudice score from each of the Spanish protocols on the Projective Test of Racial Attitudes, and compute the degree of relationship between it and the bilingual background score of the Hoffman Bilingual Schedule.

It was noted that the rectilinear relationship existing between bilingual background and racial prejudice increased progressively with age. Curvilinear analysis demonstrated that, at the four-year level, decreasing progressively through the eighth and twelfth years, there was a tendency for those subjects with least prejudice to have the highest and lowest bilingual scores.

Though this curvilinear function at the four-year level was possibly reflective of parental disposition, it indicated, nevertheless, that a profound knowledge of the Anglo culture or no knowledge of it yielded least bias toward it. Possibly, with less bilingual background, the Spanish become part of the Anglo culture and in many cases identify themselves with it. This tendency might explain the finding for many of the younger Spanish subjects to be prejudiced in favor of the Anglo.

REFERENCES

1) Helen Sargent. "Projective Methods," *Psychological Bulletin*, Vol. 42, p. 266.

2) Loc. cit.

3) Seth Arsenian. "Bilingualism in the Post-War World," *Psychological Bulletin*, Vol. 42, February, 1945, p. 70.

4) Arsenian, p. 71.

5) Arsenian, p. 71.

6) Arsenian, pp. 71–72.

INTRAUNIVERSITY VARIATIONS IN GRADING PRACTICES

BERNARD M. BASS

Louisiana State University

There has been an increasing interest among psychologists concerning criterion relevance and reliability. College grades are a criterion for many test validation studies as well as the basis for evaluating 'How to Study' courses and educational counseling programs. Grades also serve as predictors of success in graduate schools, professional schools and industry. Yet, relatively little information is available concerning the extent to which non-achievement factors contaminate the assignment of college grades.

The present study involved a three-component analysis of variance of mean grades[1] assigned by thirty-three departments[2] at three course levels[3] for four continuous semesters at a large state university. This analysis held constant the number of grades contributing to a given mean by treating the means only. Contributing to the formation of the three hundred ninety-six means were 139,659 grades assigned in the four semesters from the Fall of 1947 to the Spring of 1949.

Table 1 shows the analysis of variance of the three hundred ninety-six mean grades and the appropriate F ratios obtained.

The null hypothesis that no interaction existed could be rejected at the one per cent level of confidence. It was inferred that at different course levels, different departments deviated significantly in mean grade assigned from the course level average for all departments. For example, two departments actually graded more severely, in the absolute sense, in their junior-senior courses than in their freshman-sophomore courses while two departments assigned lower grades in graduate courses than in junior-senior courses.[4] Only seven of the thirty-three departments consistently graded lower than the level averages at all three levels and only six departments graded higher than the level average at all three levels. Thus only thirteen departments of the thirty-three studied seemed to maintain the same relative intradepartmental evaluation standards at all levels of instruction.

---

[1] Mean grades were obtained by assigning four points to each A awarded by a department at a given level, three points to each B, two points to each C, one point to each D and no points to each F awarded.

[2] Only departments giving courses at all academic levels and teaching at least five hundred students each semester were selected for this analysis.

[3] Freshman-sophomore, junior-senior and graduate levels.

[4] See below concerning the average increase in mean grades at successive levels.

TABLE 1.—ANALYSIS OF VARIANCE OF THE GRADING SYSTEM

Source of Variance	SS	d.f.	V
Among departments (d)	17.1741	32	5.367
Among course levels (l)	73.7550	2	36.878
Within cells (wc)	9.3982	297	.032
Interaction (dxi)	16.0367	64	.251
Total	116.3640	395	

$$F = \frac{V_{dxi}}{V_{wc}} = 7.930 \qquad P < .01$$

$$F = \frac{V_d}{V_{dxi}} = 2.141 \qquad P < .01$$

$$F = \frac{V_l}{V_{dxi}} = 147.156 \qquad P < .01$$

The variance associated with changes in course levels was significantly greater than the interaction at the one per cent level of confidence. The mean for all graduate courses was 3.48 or B+, the mean for junior-senior courses was 2.87 and the mean for the freshman-sophomore level was 2.43. It was inferred that the mean grade change from level to level was significant above and beyond changes due to departmental variations and semester fluctuations.

The null hypothesis that the variance of means among departments could be accounted for by the interaction variance was rejected at the one per cent level of confidence. It was inferred that above and beyond semester fluctuations and variance associated with increasing course levels, departments had differing mean grades. For example, three departments' freshman-sophomore averages were higher than a fourth department's graduate average. The range of mean grades of departments as a whole was from 2.46 to 3.42.

Assuming that the grading practices at the one university studied are common to many others, the results of this study suggest to the writer that perhaps those working on further refinements of scholastic aptitude tests may increase the efficiency of the tests in the future by developing series of regression equations specific to individual departments and course levels.

In the evaluation of 'How to Study' courses there is a need for more detailed matching of experimental and control groups. The greater mean increase in grades of the experimental group may accrue from the interaction between departments and levels. More of the experimental than the control group may take courses in departments whose grading practices become much more lenient at succeeding levels. Department and level should be included in any matching that is done.

This study also suggests the need to consider the possibilities of revising present college grading practices.

NOTES ON THE COEFFICIENT OF AGREEMENT

J. WILLIAM DUNLAP

School of Public Health, Harvard University

On numerous occasions situations arise where a number of individuals are required to rate or rank a group of items, objects, or individuals. Distributions of the ratings assigned the several items, objects, or individuals reveal varying degrees of spread, representing heterogeneity or disagreement among the raters. The degree of failure to agree can be estimated by inspecting the distributions. The disagreement can be quantified by measuring the deviation about the mean rating. Two standard statistics are available for this purpose, the mean deviation and the standard deviation. The first of these weighs the deviations linearly, and the second, by a function of the sums of the squared deviations. The latter function has the advantage of being a more elegant statistic, but is more difficult to compute. In this respect, the mean deviation is simpler, and it has the further advantage that when dealing with ratings which are of unknown reliability, of not giving undue weight to extreme deviations.

A situation which often arises is one where the research worker in dealing with the laity needs to express the degree of agreement as a percentage function to facilitate comparisons. There is no readily available function for so doing, but such a measure can be derived from either the mean deviation or standard deviation. The former has been selected here, because of simplicity of calculation and the nature of the material to which it is to be applied.

The mean deviation is the sum of the deviations about the mean, without regard to sign, divided by the number of observations. The magnitude of the mean deviation is strictly dependent on the sum of the deviations. In the case where N individuals rank a given item or person, the sum of the deviations can be readily obtained by adding the $N/2$ largest rankings and subtracting from this sum the total of the $N/2$ lowest rankings. The maximum disagreement between two raters occurs when one rater places item or person first and the second places it last. Therefore, the maximum disagreement within a group of raters occurs when half rate an item first and half rate the same item last. That is disagreement is a maximum when it equals

$$N/2(H\text{–}L). \tag{1}$$

Where: $(H\text{–}L)$ is the greatest possible range between rankings, i.e., the highest possible rank minus the lowest possible rank.

A measure of disagreement is provided by the following expression:

$$\frac{\Sigma T - \Sigma B}{N/2(H\text{–}L)} \tag{2}$$

Where: T is the top fifty per cent of the rankings and B is the bottom fifty per cent of the rankings.

Since this expression is a measure of disagreement, it follows that when disagreement is greatest the result equals one, and when the raters are in perfect agreement the result is zero. This can be adjusted so that we have an expression of agreement expressed as a percentage, as shown below:

$$\text{Coef. of Agreement} = 100\left[1 - \frac{\Sigma T - \Sigma B^{*}}{N/2(H\text{–}L)}\right] \tag{3}$$

An example will clarify the application of the formula:

Rank	Item A	Item B
1	//	
2	//	
3	/	/
4	/	
5		~~////~~

For Item A the three lowest ranks, given by the six raters are 1, 1, and 2, the sum of which is four, and the three highest ranks are 2, 3, and 4 which total to nine. The greatest possible range $(H\text{–}L)$ is 5 minus 1 which equals four and N equals 6. Therefore, substituting in (3) we have the C.A.

$$\text{C.A.} = 100\left[1 - \frac{9 - 4}{6/2(5\text{–}1)}\right] = 58\%$$

For Item B the three lowest ranks are 3, 5 and 5, totalling to

---

[*] Note that the differences in both numerator and denominator can be treated as pure numbers with regard to sign.

thirteen, and the three highest ranks are 5, 5 and 5 which total to fifteen. The greatest possible range is four. Substitution yields the following:

$$\text{C.A.} = 100\left[1 - \frac{15 - 13}{6/2(5-1)}\right] = 83\%$$

In conclusion it should be emphasized that Coefficient of Agreement is a pragmatic but practical function, useful in working with business and other personnel not versed in the intricacies of statistics. Unfortunately no sampling error is available for the function nor is the writer competent to provide one.

BOOK REVIEWS

DOUGLAS H. FRYER AND EDWIN R. HENRY, Editors. *Handbook of Applied Psychology.* New York: Rhinehart & Co., 1950. Vol. 1, pp. 380. Vol. 2, pp. 462.

These volumes contain five units dealing with the application of psychology (1) to group and individual efficiency and adjustment; (2) to business and industry; (3) to education; (4) to clinical problems; and (5) as a profession. For each of these, at the end of the second volume, a bibliography is provided listing major publications in the area as conceived by the contributors. At the beginning of Volume One, there is a three-page list of landmarks of professional psychology beginning with 1860 (introduction of psychophysical methods by Fechner) and ending in 1939 (publication of the *Dictionary of Occupational Titles*). In all, there are one hundred fifteen sections grouped into eighteen chapters.

How should a handbook be judged? Validity of organization is one criterion. On this ground, the five major sections of the handbook seem equitable. Another criterion is the allotment of space to individual topics within each major heading—here we find the articles of roughly equal length, three to five pages, with some exceptions that seem justified in the light of how much more we know in these fields. What about the stature of the contributors—are they well known, not without some eminence in the fields on which they write? Here we find a verdict somewhere between superior and excellent. To have Remmers writing on attitudes, Tinker and Paterson on typography and legibility in reading, Shartle on job analysis, Richardson on effectiveness of selection devices, Traxler on school classification, John Anderson on child development, Seidenfeld on physical disease—to pick only a few almost at random—is to show that the editors have the felicity of judgment and the persuasiveness in getting good men to work that are indispensable for the success of such an enterprise.

What about mechanical setup? The table of contents is complete and clear, and so is the alphabetical list of contributors. The typography is pleasant, dignified, appropriate. The index of contributors is of dubious value except, perhaps, for the contributors themselves. The subject index ought to err on the side of overcompleteness; we are willing to forgive uneven generality of categories if we can generally find whatever we want. But it is not hard to put this index to tests it cannot pass. For example, the entry under "Item Analysis" refers solely to the section by Wherry on item analysis and tells us nothing about the excellent page that Flanagan presents, under "Educational Test Construction," on the selection of test items. Where would one look if one wanted to read about "homogeneous grouping," or "ability grouping," or "grouping" under any name? There is nothing in the index under these titles, and the major section on these matters, "School Classification" by Traxler, would never be found through the index if one looked under "Classification." Similarly, there are additional treatments of school classification by Gordon V. Hendrickson and by Carter V. Good, who discuss instruction in elementary and secondary schools, and in college and university, respectively. No one could ever discover this by looking into the index.

What about the adequacy of references to the literature? After all, a handbook can give us merely an entering wedge for serious work. To get more deeply into a problem, we need to go to the sources the contributors cite. Here we find, perhaps, the major shortcoming of these volumes. For some reason, they give us a "selected and classified bibliography in applied psychology" which omits most of the references by the authors of the various sections. You read that "Interest studies also have considered the correlation between degree of interest in a college course and the mark received (Fryer, 1927; Meenes, 1935; Kulp, 1935) and the effect of teacher popularity upon interest in school subjects (Corey and Berry, 1938)." You may immediately want to have the title and full reference to the respective publications of Fryer, Meenes, Kulp, and Corey and Berry. If you turn to the section on educational psychology of the bibliography at the end of Volume 2, you will find not one of these. Perhaps, if you look at the reading references at the end of the section, you find bibliographies, in the more general works listed, that will lead you to the relevant publications of Fryer and the others. But probably not. One wonders then what the purpose of the fifty-nine pages of bibliography at the end of Volume 2 may be. We can not find in them what the authors of the various sections refer to. Are they perhaps a selective reading list in which the authors of textbooks and the organizers of courses may find the most significant materials in a given field? But the bibliography is too extensive for that. The irritated reader can only conclude that in designing the bibliography the editors, trying to save space, select, and organize, "o'erleaped themselves and fell on t'other side."

How successfully is duplication and overlapping avoided? Is there cross-referencing where necessary? Here again, our judgment cannot be favorable. For example, the editors allow both Hendrickson and Good to devote precious space to the problems of homogeneous grouping when these were already thoroughly accounted for by Traxler. On educational acceleration, why not substitute a cross-reference from Good's article to Pressey's and so save another half-page? Similarly, nothing is said in the section on exceptional children that will take the reader into the two chapters on clinical psychology dealing with mental deficiency, community child guidance clinics, speech difficulties, and the like.

What is the picture of applied psychology that one gets from this handbook? One impression is that of tremendous diversity. Applied psychologists have not feared to tread. Psychological warfare, ventilation, metropolitan police personnel, building illumination, railroads, radio and television research, visual efficiency, physical disease, the courts, prisons and penitentiary, the rehabilitation of the aged, librarianship, eugenics, the theater—these titles of sections suggest the range of human concerns and economic activities that applied psychologists feel qualified to deal with. Among the few areas untouched not only by the editors and their authors, but also by applied psychology itself, are agricultural work and politics. But from these volumes, one gathers courage to predict that, before many years, the applied psychologists will have sent out scouts, deployed their forces, and colonized these areas.

As is perhaps appropriate in writings on applied psychology, we get little of the systematic and theoretical flavor that permeates the efforts of psychologists who do not call themselves 'applied.' The assumptions, conceptualizations, constructs, theoretical frameworks, and chains of hypotheses that characterize psychology in general are here all right, but they are never made explicit and systematic. As we all know, the applied psychologist is out to solve practical problems. That this can be done, as some action researchers insist, without neglecting the task of building systematic theory, would never be guessed from this handbook. Rather, we have a kind of 'interdisciplinary' approach in that many of the articles are concerned not so much with psychology as with the same practical, horsesense kind of intelligent attack on problems that any bright layman might undertake, especially if given a smattering of the introductory course and a few statistical techniques.

For readers of this JOURNAL, we consider in detail only the two chapters on educational psychology, which fill eighty-five of the large double-columned pages.

Chapter IX, Educational Psychology: Techniques of Educational Adjustment, begins with a section by Gertrude Hildreth describing the school psychologist's services: coördinating instruction, studying individual children, and dealing with the slow learner, the behavior problem child, failures in school subjects, reading disabilities, physical handicaps, and gifted children. She also outlines the psychological testing, observational, and interview techniques of school psychologists and their functions in promoting pupil-teacher relationships, curriculum adjustment, cumulative records, and home-school relations.

Arthur Traxler gives a circumspect analysis of the considerations involved in the classification of students into different grade levels and within single grades. Chronological age, age grade, mental age, IQ, reading, and achievement age, as well as various formulas, are described and succinctly evaluated as bases for the grade classification of pupils. Ability grouping within single grades, inter-age grouping, and automatic promotion are similarly treated. His discussion ends with a cogent advocacy of modern measurement and recording procedures integrated into a well organized guidance program as the way out of the dilemmas confronted in any more mechanical scheme of pupil classification. Similarly, adequate individualization of educational experiences depends in the final analysis on the ability, initiative, and imagination of the classroom teacher, above and beyond any classification scheme.

Percival Symonds' section on principles of child guidance begins with a distinction between preventive and corrective guidance and points to the uniqueness of every case as militating against the application of general rules. After a rapid summary of findings on the criteria of treatability, he discusses methods of gaining the child's coöperation and the long-term nature of most readjustment processes. Since therapy is more and more considered a 'relationship,' attention to both the child and his parents is ordinarily called for. But different clinicians should work with different members of the family. The therapy thus undertaken can be aimed at the elimination of symptoms as in direct conditioning processes for curing enuresis; at removing frustrations as in changing a child's teacher or interesting him in a recreational group; or at the release of tension, change of attitude, and development of self-understanding as in psychotherapeutic interviews. As to psychotherapy, Symonds quotes David Levy's *Maternal Overprotection* with its discouraging findings concerning psychotherapy as compared with education, environmental change, advice, and demonstration. A brief statement on criteria of improvement, mainly concerned with the dangers of too early cessation of treatment, closes the section.

Morton Seidenfeld's section on the care and treatment of language deficiency describes the need for, organization, and functioning of the special training units set up by the Army in World War II to teach basic reading, writing, and arithmetic skills to the thousands of 'illiterate' young men inducted by selective service. After a straight-forward exposition of how the program operated, the author draws the lessons for civilian life that such skills are best taught to adults when built into the learner's life milieu, that utilitarian values must be stressed, that relatively large gains in knowledge should be sought quickly at the start to enhance motivation, and that the social problems in removing illiteracy are at least as worthy of attention as the educational ones.

T. Ernest Newland and Lester N. Myer, dealing with exceptional children, distinguish between the education and care of such children, offer a general definition of 'exceptional children,' and then briefly discuss the nature and incidence of mentally retarded, mentally superior, visually handicapped, acoustically handicapped, speech handicapped, and physically delicate and orthopedic cases. Their estimates of the percentages of American children in each of these groups are three, three, twenty-five hundredths, three, six, and two, respectively. The educational program for such children should deal with social and emotional adjustment as well as the three r's. Only about fifteen per cent of the cases are presently cared for in special classes and schools; the other eighty-five per cent (almost two and a quarter million) are either in regular classes or unlocated. Psychological services deal with the discovery and examination of exceptional children and with making recommendations for their educational adjustment, especially their social and personality training. Urgent needs exist for publicity among educators concerning this kind of education, for the early examination and study of all children, and for research on the effectiveness of special education programs. Finally, the authors answer the questions, Why educate exceptional children? Why expend extra effort and expense in their interest?—in terms of humanitarian values both to the children helped and to society.

John C. Flanagan's article on educational test construction goes into the following topics: (1) basic principles of test construction, in which validity and 'efficiency' are defined; (2) outlines and specifications for tests; (3) writing and editing test items, in which four levels (facts, generalizations, applications, and significances) are described and illustrated; (4) the selection of test items, where he considers item validity (discrimination), difficulty, inter-correlation, time required, objectivity, face validity (acceptability), and answer defensibility; (5) the organization of a test, logically and typographically; (6) equating the forms of a test; (7) the scaling of raw score units; (8) techniques for scoring test items; (9) the provision of norms, in which he outlines five primary types of interpretation: absolute performance in terms of content, relative performance in terms of rank in a specific group, performance with respect to that of the average individual in various specified groups, performance with respect to the individual's own previous performance, performance with respect to the individual's performance in other fields; and (10) other factors in the interpretation of results (motivation, guessing or scoring errors, and accuracy).

Sidney L. Pressey's section on accelerated higher education defines acceleration as "moving through an educational program in a shorter time or at earlier stages than is usual," and then raises two questions: What are the most desirable ages for a given educational program? and What are the optimum rates of progress? His answers are based on historical precedent and on the findings of experimental studies. The general conclusion is that acceleration has relatively satisfactory outcomes and that the major problem is how to overcome the "inertia of the enormous total institution of American education, which has made twelve years of pre-college and four years of college, and the usual school calendar, a basic pattern for student life and faculty habits of thinking . . . "

Chapter X, Fields of Educational Psychology, begins with a section on child development by John Anderson. After a brief historical introduction, Anderson delineates the interdisciplinary character of child study and the interplay of science and practice that have characterized the field. Research on young children has considered the child as the independent variable, studied the effects of environment on behavior, and examined behavior *in situ* through various observational techniques. Education of young children in nursery schools, training of professional workers, relationships of child development to general education, parent education, popular writing and radio, and clinical work with young children are the additional topics briefly characterized.

Gordon Hendrickson's section on instruction in elementary and secondary schools takes up generalizations derived from research on individual differences, readiness to learn, motivation, general patterns of teaching, audio-visual aids, and transfer of training. After pointing out the facts of intra-individual and intra-class variability, he describes our experience with ability grouping and individualized instruction and draws the conclusion that teachers should expect different levels of achievement from their pupils unless the goals are severely circumscribed. Under readiness to learn, we find the problem outlined with special attention to reading and algebra; the conclusion is that readiness depends not only on characteristics of the pupil but on the teacher's skills and other factors. Under motivation are considered the inadequacy of instinct doctrine and the findings from studies of praise vs. reproof, reward and punishment, knowledge of results, and individual vs. group competition; such studies too often neglect the learner's understanding of the learning situation. As to general patterns of teaching, the ones considered are the recitation and laboratory methods, projects and activities units, and the unit-mastery plan, in each case with a summary of evaluative research on the method; the general implication is that teaching success depends more on the adaptability of the teacher than on any formal method. On audio-visual aids, we find condensed summaries of the research on the effectiveness of motion picture, radio and recordings; the general limitation of these devices in enforcing pupil passivity is pointed out. Transfer of training is treated first with a review of pertinent research of the more traditional kind and then, with more encouraging results, in terms of the carry-over from activity-unit and other newer type procedures; studies of the latter kind, properly considered relevant to transfer of training, have yielded encouraging support for those who seek positive transfer from classroom activities to intellectual, emotional, and social maturity. A final meaty paragraph gives us principles useful in securing maximum transfer.

Instruction in college and university is treated in a section by Carter V. Good. The topics he takes up are techniques of investigation, class size, ability grouping and acceleration, motivation, instructional methods, audio-visual aids, study procedures, procedures for adjustment of the inferior and the superior student, and examinations and marking. For each topic, surveys of the findings of representative investigations are given. From all this, Good sees needs for the evaluation in higher education of trends toward adjustment of the program to the individual student, toward comprehensive achievement examinations as substitutes for time spent in granting academic credit, toward organization of broader instructional divisions such as biological sciences or social sciences, and toward programs of general education.

The section on predicting success in college and university by Edwin R. Henry begins with the rationale of admissions programs and of the problem of selective admission, discusses criteria of success in college, and outlines predictive technique in terms of the well-known regression equation. The remainder of the section takes up the typical findings with such predictors of academic success as secondary-school record, secondary-school courses, college aptitude tests, personality measures, age at admission, size of family, part-time employment, and paternal occupation.

Horace B. English and George Nagy begin their section on "Professional School Admissions" by contrasting professional with liberal education in terms of the former's importance to the individual and society, the personality traits and abilities required, and the greater need for ethical, social service motivations. Criteria of professional success are continuance in the profession, membership in professional associations, advancement, research and publication, reputation, and income. Among the selective procedures commonly used are academic record; achievement, aptitude, interest and personality measures; health, financial, sex, race, religious, and residential characteristics. Following this general survey, the authors summarize the practices and results that have been obtained in graduate schools, medicine, dentistry, nursing, theology, law, engineering, and teaching. The conclusion is that the best single predictor of academic success in professional training is previous academic grades; when combined with standard achievement tests, predictive accuracy reaches the limit possible with imperfectly reliable criteria. The greatest need in the field is to develop criteria less narrowly academic and more concerned with ethics, social outlook, character, drive, and personality.

William D. Glenn, Jr. and Wallace Gobetz, writing on "College Student Personnel Services," outline the student personnel point of view and the wide variety of jobs that are needed to perform student personnel functions on present-day campuses. The authors then give brief characterizations of the problems, findings, and procedures of student personnel workers in dealing with admissions, selection of curriculum, orientation, health, extra-curricular activities, counseling, remedial work, financial assistance, placement and follow-up. A final paragraph on needed research points out the inadequacies of previous work and a host of problems that have been relatively untouched.

Outside of these two chapters on educational psychology, there are many sections in other chapters of interest to the educational psychologist. Among the most pertinent are sections on motivation and self-direction (Douglas H. Fryer), interest in work and play (Donald E. Super and Jack W. Dunlap), aptitude and intelligence (George K. Bennett), guided learning (T. R. McConnell and G. Lester Anderson), vocational and educational satisfaction (Robert Hoppock and Donald E. Super), administration of psychological services in public schools (Bertha M. Luckey), state administration in public schools (Warren W. Coxe), and administration of psychological services in college and universities (William D. Glenn, Jr., and Wallace Gobetz).

The section on guided learning by McConnell and Anderson defines learning, discusses its ubiquity and then outlines the process in terms of the eleven characteristics listed by Dashiell in 1935. The function of motivation, the variety of changes in behavior that learning may bring about, and transfer of training are sketched. Guidance in learning stresses the need for more than mere repetition—the importance of consequences or effects, the proper distribution of practice, difficulty, and organization. In learning an active skill, we need first to define its nature, and then by direct observation, or manual guidance, properly organized as to wholes or parts, in a situation maximally similar to that in which the skill will be used, to achieve a flexible pattern adaptable to the perceived demands of the situation. Under ideational learning, the authors stress the importance of meaningfulness and organization as perceived by the learner and then take up whole and part learning, practice techniques, and the nature of meaningfullness in learning. Forgetting is considered as the result of retroactive inhibition. Limitations to learning are characteristics of the learner such as his ability, various organic states, and physical and social surroundings. In general, the principles of learning are not invariant laws but rather expressions of significant considerations in the control of learning.

All in all, the editors and publishers of these volumes have done applied psychology a distinct service. These volumes may well become, if not the *vade mecum* because they are too bulky, at least a frequent source of quick acquaintanceships with dozens of areas. Graduate students, personnel men, thesis directors, and workers in the vineyards of applied psychology will find help here. And not least, the historian of students of industrial society will be fascinated by this handbook a century from now.

N. L. GAGE

University of Illinois

REGINA MARY GOFF. *Problems and Emotional Difficulties of Negro Children.* New York: Bureau of Publications, Teachers College, Columbia University, 1949, pp. 93.

This study investigates the kinds of difficulties Negro children ages ten to twelve report that they experience because of their race. Ninety children in New York City and sixty in St. Louis interviewed by a Negro investigator from January to July, 1946, reported an average of three difficulties. These were categorized into overt manifestations (direct ridicule, physical ill-treatment, and aggressive behavior) and subtle manifestations (discrimination, rude treatment, and indirect disparagement). Children were asked how they felt during these experiences, what impulses were aroused, what action they took, and what advice they received from parents. One hundred and fifty mothers (not necessarily mothers of children interviewed) were requested to relate difficulties on account of race reported to them by their children, the guidance given, and its effectiveness.

Comparisons were made of data by city, socio-economic level, and sex. A similarity in basic pattern of difficulties was found in New York and St. Louis although in New York subtle, indirect manifestations of prejudice and rude treatment were more common; and in St. Louis, the overt forms of discrimination. The most frequently reported difficulty was ridicule; next in rank was disparagement created by movie and radio stereotypes. Distribution of difficulties had negligible relationship to socio-economic level; however, low income children in exclusively Negro neighborhoods pay relatively greater attention to implied prejudice than do upper income children. Data suggest that girls are more sensitive than boys to manifestations of prejudice.

Ridicule and physical aggression arouse resentment and impulse to fight; discrimination and rude treatment arouse inferiority feeling. Relatively more low income children than upper income children report inferiority feeling aroused by movie and radio stereotypes. Indirect disparagement provoked resentment accompanied by an inclination to do nothing about it. Low income children exhibited no greater aggressiveness in response to race difficulties than upper class children. The author explains this unexpected finding on the basis of selective sampling. Low income children were obtained through listings of community centers which are probably attended by less aggressive children.

Parental guidance appeared to be ineffective. Author stresses need for parent education among Negroes. Instructions to withdraw were most frequently given by parents. They felt that churches were ineffective in improving race relations, were skeptical of intergroup organizations, and favored more Negro teachers for Negro children.

Three illustrative interviews are presented. After reviewing her data the author concludes that Negro children desire to strike back in prejudice situations, but suppress their emotions and withdraw instead of taking active steps to face their problems. In the course of her discussion of the implications of her study she states: "It is apparent that the Negro child needs an enriched program of training which places more emphasis on the building of attitudes toward himself—attitudes especially of self-esteem, self-respect, and self-confidence."

The difficulties and pitfalls of the interview method as a technique of securing information must be considered in evaluating the findings of this study. Some of these are recognized and described by the author.

This exploration into an area of paramount importance for American society and particularly for education is commendable. It extends to a very narrow age range and generalizations made may give misleading impressions if this limitation is overlooked.

PHILIP M. KITAY

Adelphi College

ALGERNON COLEMAN, CLARA B. KING, CLARE BALLUFF, AND ROBERT H. FIFE. *An Analytical Bibliography of Modern Language Teaching*, Vol. III, 1937–1942. New York: King's Crown Press, Columbia University, 1949, pp. 549.

This volume continues the series of reports begun in 1928. It brings up to 1942 the lists of books, articles, reports, and bulletins concerned with progress in modern language teaching. Foreign language teaching abroad is omitted from the present volume. Each of the eight hundred fifty-three items included presents bibliographical material, a summary of contents and an evaluation of the contribution.

Some of the trends evident during the period covered are: an increase of emphasis upon the relation of modern languages to social sciences and their integration with cultural studies; a vigorous interest in achievement tests; appreciation of experimental results; a rising interest in Spanish and Portuguese;

increased emphasis upon oral training and an increase of public interest in foreign languages. The more extensive sections deal with (1) aims, materials and method, (2) curricula-syllabi, and (3) examinations and tests.

The summaries and evaluations are well done. This over-all view of trends in modern language teaching should be very useful to workers in the field.

MILES A. TINKER

University of Minnesota

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 42 November, 1951 Number 7

A PROBLEM IN FACTORIAL DESIGN: THE DEVELOPMENT OF A PICTURE AND INSET TEST FOR DETERMINING YOUNG CHILDREN'S RESPONSES TO PERSONS OF DIFFERENT SKIN COLOR*

CATHERINE LANDRETH and ELIZABETH FRAZIER PLATT
University of California

Recent investigations of young children's social perception, in particular their perception of racial characteristics,[1,2,3,4,5,6,7,8] direct attention to the need for effective solution of the methodological problems involved in such studies. The following report of an investigation, currently in progress, presents the writers' solutions to the problems they encountered in undertaking a study of young children's responses to persons of different skin color.

Investigations to date of this aspect of young children's behavior have been concerned with children's racial identification and consciousness of self[1,2,3,6] with the relative significance of race, sex and facial expression on the choice of playmates[5] and with children's perceptions of the social rôles of Negroes and whites.[7,8]

While all of these are fundamental problems, a glance at Table I summarizing information on the samples reveals sampling procedures that qualify ones acceptance of some of the generalizations on the population groups represented. For example in two reports[4,6] numbers are so small as to make generalizations impossible save for the twenty odd children concerned. In four,

---

* Acknowledgment is due to graduate students Phyllis Du Pont, Willa H. Schmidt, Gloria Benson and Barbara C. Johnson for preliminary group studies, to Dr. Stanley Nash for mathematical assistance, and to Dr. Rheem Jarrett and Dr. Harold E. Jones for reading the manuscript.

though samples are large for the total study, comparisons for subgroups representing skin color, race, socio-economic status and segregation are made between such disparate numbers of children as nineteen, twenty-one and one hundred,[3] twenty-four, eighty-four and twenty-seven,[5] ninety and one hundred fifty-two[7] and eighty per cent and twenty per cent of the sample.[5] Three of the reports are concerned solely with negro children.[1,2,3] Two[7,8] are limited to children of five years and over. Measures of significance for differences obtained are lacking completely in three reports.[3,4,6]

PURPOSE OF INVESTIGATION

In planning the present investigation it was, therefore, decided to study responses to persons of different skin color by equal numbers of children—of both sexes— of three and five years of age—of upper and lower socio-economic status—of the Negro and white race—living in a Northern and a Southern State.

Table II summarizes facts concerning the sample. The groups chosen for study represent readily distinguishable groups available in the population. The ages three and five were selected with the thought that three-year-olds would be the youngest group from whom coöperation might be obtained. A study of four-year-olds' responses is contingent on the difference in findings between the three- and five-year-olds. At present data are still being gathered on these groups of children in California.

Socio-economic status was not made a basis of distinction for Negro children because of the difficulty encountered in establishing criteria with the limited assistance available through the schools. It is, however, the investigators' hope to study this factor later.

While a comparison of responses of children in mixed groups (Negro and white) with those in segregated groups would be interesting, it was found that mixed groups do not exist in the local school system in Berkeley and Oakland. Housing area restrictions appear to be the reason for white or Negro schools with a negligible percentage of children of the other race in attendance.

A glance at Table II indicates the type and number of intergroup comparisons possible with the experimental design indicated by the sample.

TABLE I.—SAMPLES USED IN REPORTS OF YOUNG CHILDREN'S REACTIONS TO DIFFERENCES IN SKIN COLOR

Reports	Number Cases	Age	Sex	Race	Socio-Ec. Status	Group Exp.	Skin Color
1	150	50 3 yr. 50 4 yr. 50 5 yr.	75 male 75 female	150 Negroes	5 W.P.A.N.S. 1 Pvt. N.S. 1 Pub. Sch. Nos.	Segregated	30 light 66 medium 54 dark
2		Same as in No. I					
3	140	3 yr. 4 yr.		140 Negroes		21 Mixed 19 Semi-segregated 100 Segregated	
4	27	2–9 to 4–4		15 Negroes 12 White			
5	135	Pre-sch. Kind.	57% 43%	20% Negroes 80% White	24 White high 84 White low 27 Negro low	24 White segregated 28 Negro segregated 73 Mixed	
6	24	2–3 to 5–1	16 male 8 female	7 Negro 17 White	W.P.A.N.S.		
7	242	Kind. First Second Grade		90 Negro 152 White	Middle and Lower		
8	250	Kind. First Second Grade			Middle and Lower		

PICTURE INSET TEST

In developing a test situation it was decided that a response involving interesting activity would result in readier coöperation from the three-year-olds than one calling for verbalization. A picture with one figure which could be completed by adding an inset containing another figure, offered such a possibility.

Table II.—Young Children's Responses to Persons of Different Skin Color
Population Sample
N-576

Age	3 years			5 years		
Race	Negro		White	Negro		White
Soc. Ec. Status	Undefined	Upper	Lower	Undefined	Upper	Lower
Northern State (California)						
Girls	24	24	24	24	24	24
Boys	24	24	24	24	24	24
Southern State (Alabama)						
Girls	24	24	24	24	24	24
Boys	24	24	24	24	24	24

Two series of pictures and insets were therefore devised, one for boys and one for girls. Each series contains eighteen pictures. As Table III illustrates, in six pictures the picture figure is white; in six, brown; in six, black. In four out of each six of these pictures, the picture figure is a child of the same sex as the child to be tested; in one out of each six the picture figure is a child of the opposite sex; and in one, a woman.

The figure represented in the accompanying insets is either a child of the same sex as the child in the picture, a baby, a woman or a man. Each picture has three accompanying insets

TABLE III.—YOUNG CHILDREN'S RESPONSES TO PERSONS OF DIFFERENT SKIN COLOR. PICTURE INSET SERIES FOR BOYS

Picture Figure		Inset Figure			
Skin Color	White	Skin Color	White	Brown	Black
1A*	Boy	Boy	a†	b†	c†
2A	”	Baby	a	b	c
3A	”	Woman	a	b	c
4A	”	Man	a	b	c
5A	Woman	Man	a	b	c
6A	Girl	Girl	a	b	c
Skin Color	Brown	Skin Color	White	Brown	Black
1B*	Boy	Boy	a	b	c
2B	”	Baby	a	b	c
3B	”	Woman	a	b	c
4B	”	Man	a	b	c
5B	Woman	Man	a	b	c
6B	Girl	Girl	a	b	c
Skin Color	Black	Skin Color	White	Brown	Black
1C*	Boy	Boy	a	b	c
2C	”	Baby	a	b	c
3C	”	Woman	a	b	c
4C	”	Man	a	b	c
5C	Woman	Man	a	b	c
6C	Girl	Girl	a	b	c

* ABC refers to skin color of figure in picture.

† abc refers to skin color of figure in inset.

Note: Figures 1 to 6 refer to picture and inset person-combination, thus 3A—white boy with a woman.

which are identical in all respects except the skin color of the person represented.

Each picture when completed with an inset portrays two persons engaged in an activity with which any young child would be familiar. Activities are playing, eating, bathing, riding, walking and waking up in the morning. The pictures are designedly neutral in feeling tone in order to be within the comprehension of three-year-old children.

Construction details concerning the pictures are that they were painted with a quick drying lacquer on 12″ × 15″ masonite rectangles and finished with a clear shellac on both sides. Colors chosen were recommended by the University Decorative Art Department as being equally pleasing with white, brown and black skins. Negro facial characteristics other than skin color were suggested but not emphasized. Slight variations in background and clothing were made for the six A, B, and C pictures and insets so that children might not tire of them. For instance, IA shows two boys piling blocks in a wagon; IB, two boys on a rocking boat; and IC, two boys making a sand castle. Figure I shows two different picture and inset person-combinations. Inset figures are light white in one, black in one and brown in another inset.

Between usage pictures were stored in flannel bags hooked on to a portable storage rack.

DATA AVAILABLE FROM RESPONSES TO PICTURE INSET TEST

When the series of pictures is presented on three separate occasions with the three possible pairings of the three insets, data are offered on fifty-four choices of an inset for a picture. In thirty-six choices the child has a chance to match the skin color of the figure in the inset with that of the figure in the picture. The child also has thirty-six opportunities to register a preference for white, brown or black skin color as each inset with this skin color is presented thirty-six times as one of a pair from which he makes a choice.

In addition to obtaining data on 1) the child's matching and 2) preference choices for each skin color—white, brown and black, it is obvious that it is also possible to compare such matching and preference choices in terms of 3) the different persons represented in the inset and 4) in terms of the first and third presentation of the series.

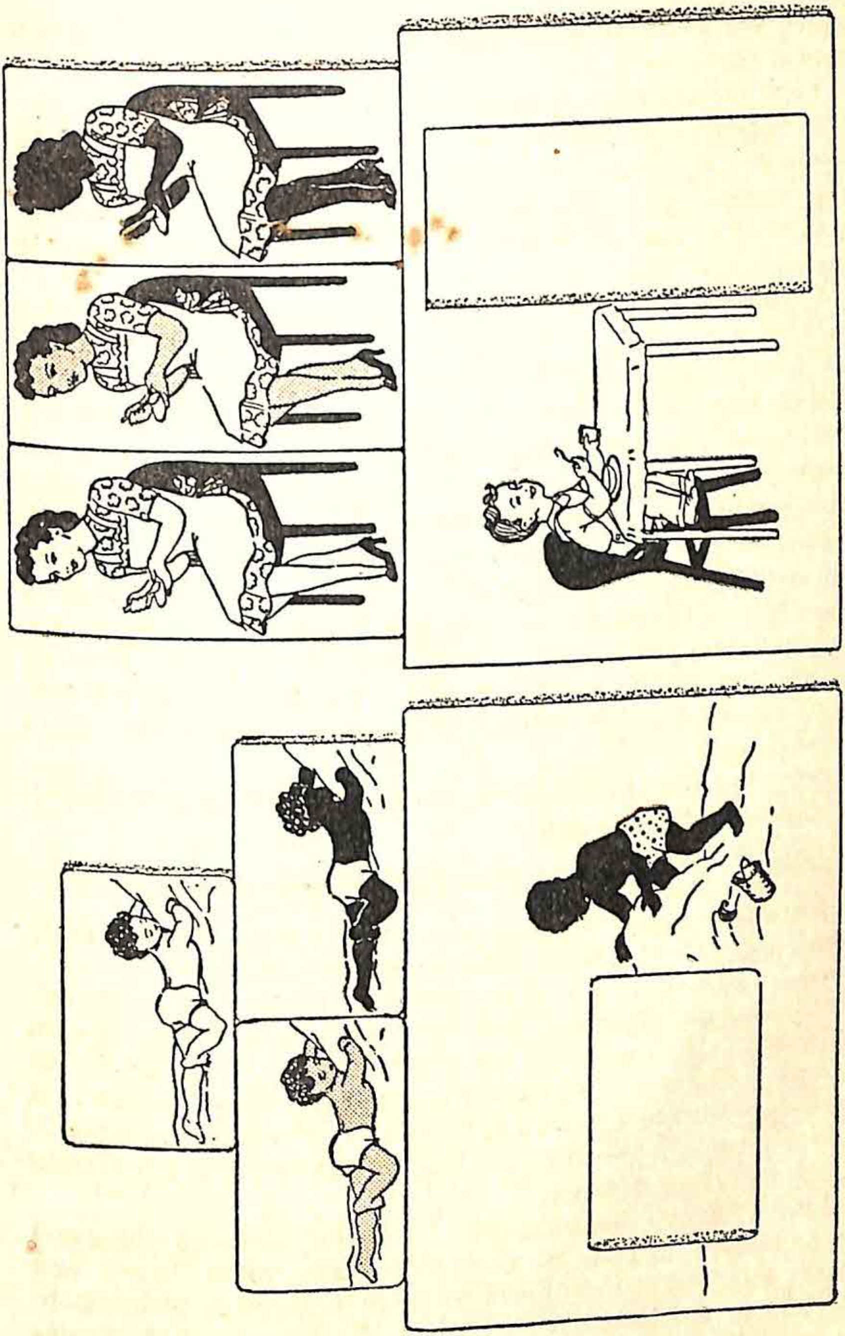


Fig. I.

Data from this latter comparison could be taken as evidence either of consistency of choice or development of a solving principle or hypothesis.

As a pair of insets necessarily assume a right and left position it is also possible 5) to compare—the number of times the child shows a preference for or matches the skin color of the figure in an inset presented to his right or his left.

If children's responses to the picture inset test are to be interpreted in terms of matching and preference reactions it is necessary that the order of presentation of the pictures does not introduce systematic variables and that it is similar for all groups of children in terms of the comparisons to be made.

PRESENTATION FACTORS HELD CONSTANT

Right-handed choices.—As a preliminary group test revealed a strong tendency of the three-year-olds to choose the inset presented to their right it was obvious that for each presentation of the series of eighteen pictures, white (a) brown (b) and black (c) inset figures must appear equally often to the child's right and to his left and must equally often match the skin color of the figure in the picture when presented to right or left.

Perseveration.—A tendency to perseverate had also to be counteracted by randomising or systemizing the order of presentation of the skin color of the figure in the picture and that of the figures in the insets.

Table IV summarizes factors held constant in presenting the insets to each group of children. Table V indicates the way in which presentation order of paired insets is developed. Many other examples besides the two presented are possible.

After working out a presentation order for the paired insets it was next necessary to randomise, or more correctly systemize the order of presentation of the skin color of the stationary figure in the picture. Systemization was in terms of having no skin color follow itself, thus:

1 A	1 B	1 C
2 C	2 A	2 B
3 B	3 C	3 A
4 A	4 B	4 C
5 B	5 C	5 A
6 C	6 A	6 B

Forty-two different presentation orders can be devised which meet this requirement.

The next problem was to combine three different picture systemization series (one for each presentation) with a set of paired insets meeting the requirements exemplified in Table V.

TABLE IV.—FACTORS HELD CONSTANT IN PRESENTATION OF 54 PAIRS OF INSETS

Presentation	Skin color of figure in inset	Number presented			Matching possibilities		
		Total	Right	Left	Total	Right	Left
First	White a	12	6	6	4	2	2
	Brown b	12	6	6	4	2	2
	Black c	12	6	6	4	2	2
Second	White a	12	6	6	4	2	2
	Brown b	12	6	6	4	2	2
	Black c	12	6	6	4	2	2
Third	White a	12	6	6	4	2	2
	Brown b	12	6	6	4	2	2
	Black c	12	6	6	4	2	2
Total	White a	36	18	18	12	6	6
	Brown b	36	18	18	12	6	6
	Black c	36	18	18	12	6	6

Eighteen matching choices possible on the right and on the left.

Equal occurrence of each permutation: ab ac bc, ba ca cb.

Each picture matched once on right, once on left.

Each picture and inset person combination presented with a white, a brown and a black skinned figure on the right and on the left in each presentation.

Table VI presents a series for one group of children, prepared as a check and tabulation sheet. This check sheet is being used with five-year-old Negro and white children in California and in Alabama.

TABLE V.—TWO EXAMPLES OF WAYS IN WHICH PAIRS OF INSETS MAY BE PRESENTED. HOLDING CONSTANT FACTORS DESCRIBED IN TABLE IV

Key I. 1 child with child inset.
2 child with baby inset.
3 child with woman inset.
4 child with man inset.
5 woman with man inset.
6 child with child inset (other sex)
Key II Number of presentation I, II, III.
Key III Right or Left R or L.
Key IV Skin color of central figure A, white; B, brown; C, black
Key V Skin color of inset figure a, white; b, brown; c, black

Picture.	I	II	III
	R-L	R-L	R-L
1A	a-b	c-a	b-c
2A	b-a	a-c	c-b
3A	a-c	c-b	b-a
4A	c-a	b-c	a-b
5A	c-b	b-a	a-c
6A	b-c	a-b	c-a
1B	c-a	b-c	a-b
2B	a-c	c-b	b-a
3B	c-b	b-a	a-c
4B	b-c	a-b	c-a
5B	b-a	a-c	c-b
6B	a-b	c-a	b-c
1C	b-c	a-b	c-a
2C	c-b	b-a	a-c
3C	b-a	a-c	c-b
4C	a-b	c-a	b-c
5C	a-c	c-b	b-a
6C	c-a	b-c	a-b
Total Presented			
a	6-6	6-6	6-6
b	6-6	6-6	6-6
c	6-6	6-6	6-6
Matching			
a	2-2	2-2	2-2
b	2-2	2-2	2-2
c	2-2	2-2	2-2
Permutations			
a-b	c-a	b-c	9 of each
b-a	a-c	c-b	9 of each

Picture.	I	II	III
	R-L	R-L	R-L
1A	a-b	b-c	c-a
2A	b-a	c-b	a-c
3A	b-c	c-a	a-b
4A	c-b	a-c	b-a
5A	c-a	a-b	b-c
6A	a-c	b-a	c-b
1B	b-c	c-a	a-b
2B	c-b	a-c	b-a
3B	c-a	a-b	b-c
4B	a-c	b-a	c-b
5B	a-b	b-c	c-a
6B	b-a	c-b	a-c
1C	c-a	a-b	b-c
2C	a-c	b-a	c-b
3C	a-b	b-c	c-a
4C	b-a	c-b	a-c
5C	b-c	c-a	a-b
6C	c-b	a-c	b-a
Total Presented			
a	6-6	6-6	6-6
b	6-6	6-6	6-6
c	6-6	6-6	6-6
Matching			
a	2-2	2-2	2-2
b	2-2	2-2	2-2
c	2-2	2-2	2-2
Permutations			
b-a	a-c	c-b	9 of each
a-b	c-a	b-c	9 of each

TABLE VI

Young Children's Responses to People of Different Skin Color

Child's Name____________________
Parent's Occ.____________________
Address____________________
School____________________
Experimenter____________________

Age____________________
Sex____________________
Skin Color____________________
School Group____________________
Teacher____________________

			Did Match			No Match			Summary					
	Picture	R L	a	b	c	a	b	c	1	2	3	4	5	6
Person	1 A	c a												
	2 C	b a												
DATE	3 B	a b												
	4 C	a c												
	5 B	c a												
	6 A	b a												
Relative	1 B	b c												
	2 A	a c												
	3 C	c a												
	4 A	c b												
	5 C	b c												
	6 B	a c												
Person	1 C	a b												
	2 B	c b												
	3 A	b c												
	4 B	b a												
	5 A	a b												
	6 C	c b												
	Total													
Relative	1 B	a b												
	2 A	c b												
DATE	3 B	c a												
	4 C	c b												
	5 B	b c												
	6 A	a c												
Person	1 C	c a												
	2 B	b a												
	3 C	b c												
	4 A	b a												
	5 C	a b												
	6 B	c b												
Relative	1 A	b c												
	2 C	a c												
	3 A	a b												
	4 B	a c												
	5 A	c a												
	6 C	b a												
	Total													
Person	1 C	b c												
	2 B	a c												
DATE	3 A	c a												
	4 B	c b												
	4 A	b c												
	6 C	a c												
Relative	1 A	a b												
	2 C	c b												
	3 B	b c												
	4 C	b a												
	5 B	a b												
	6 A	c b												
Person	1 B	c a												
	2 A	b a												
	3 C	a b												
	4 A	a c												
	5 C	c a												
	6 B	b a												
	Total													
	Totals													

Preference

1	2	3	4	5	6		
						a	
						b	
						c	

						a	
						b	
						c	

						a	
						b	
						c	

						a	
						b	
						c	

Totals

Negro Ident.

B	G
a	a
b	b
c	c

Color Ident.

1	1
2	2
3	3

Match	A						
	B						
	C						

As six groups of children are to be studied in California and in Alabama, five other sets, each of three different systemizations of picture skin colors and their accompanying inset pairs, were also prepared.

In combining picture and inset series it was found, to the investigator's chagrin, that it was not possible to assign random numbers to the forty-two picture series and the various inset series and then combine them. Even though such random combinations met all requirements listed on Tables IV and V they introduced other problems. In some cases runs of four of a kind of skin color of inset figure appeared on the right thus:

1 A	ab
2 B	cb—4 bs on one side
3 C	ab
4 A	cb

In others there were runs of matching skin colors on one side thus:

C	cb
A	ac
B	ba
C	ca
B	bc
A	ab

It was therefore arbitrarily decided to eliminate all combinations in which there was a run of four or more of a kind of skin colors on the right or the left, or five or more matching skin colors on the right or the left. It was thus that the six different check sheets, one of which is shown in Table VI were developed for the six groups in California and Alabama.

As an additional measure in overcoming the influence of systematic variables in presentation order each group of forty-eight children was divided into three subgroups of sixteen, and given the following presentation orders on the three occasions on which they were tested.*

---

* This suggestion was offered by Dr. W. G. Cochron.

Groups	Occasions		
	1	2	3
	Presentation orders		
X	I	II	III
Y	II	III	I
R	III	I	II

MANNER OF PRESENTATION

All groups in California are being tested by one white experimenter. The series of eighteen pictures, one for boys and one for girls, is presented on three occasions, a week apart. Each child is tested singly. The procedure takes from five to ten minutes for each child. The child is told, in the case of the first picture, "Here is a picture of two boys (or girls) playing. One of the boys is out of the picture. Which of these boys would you like to put in?" The insets are presented above the inset space.

At the end of the third presentation the experimenter presents six insets of a white boy, a brown girl, a black boy, a white girl, a brown boy and a black girl, and laying them in front of the child says, "Some of these children are Negroes. Would you show me all the Negro children? Put your finger on all the Negro children."

The purpose, here, is to obtain data which may be of interest in comparing children's responses to persons of different skin color in terms of their expressed awareness of its valid significance.

The experimenter then tests the child's ability to match skin color by presenting three rectangles painted with samples of skin color.

b a	c b	a c
a	b	c
a b	b a	c b
a	b	c

She points to the one in front of the child and says, "Show me the one like this. Put your finger on the one like this."

In summary the test procedure devised for determining young children's responses to persons of different skin color holds constant all presentation factors which preliminary trials indicated might affect the results. Differences in response between groups of children may therefore be interpreted in terms of native, developmental and environmental factors.

A report presenting and discussing the responses of the six groups of children tested in California is now in preparation. The six groups of children in Alabama have not yet been tested.

REFERENCE

1) Kenneth B. Clark and Mamie K. Clark. "The development of consciousness of self and the emergence of racial identification in Negro preschool children." *J. Soc. Psych.*, 10: 591–599, 1939.

2) ——— ———. "Skin color as a factor in racial identification of Negro preschool children." *J. Soc. Psych.*, 11: 159–169, 1940.

3) ——— ———. "Segregation as a factor in the racial identification of Negro pre-school children." *J. Exp. Educ.*, 8: 161–163, 1939.

4) Mary Ellen Goodman. "Evidence concerning the genesis of interracial attitudes." *American Anthropologist*, 48: 624–630, 1946.

5) Evelyn Helgerson. "The relative significance of race, sex, and facial expression in choice of playmate by the preschool child." *J. Negro Educ.*, 12: 617–622, 1943.

6) Ruth E. Horowitz. "Racial aspects of self-identification in nursery-school children." *J. Psych.*, 7: 91–99, 1939.

7) Marian J. Radke and Helen S. Trager. "Children's perceptions of the social rôles of Negroes and whites." *J. of Psychology*. 29, 3–33, 1950.

8) Marian J. Radke, Helen S. Trager and Hadassah Davis. "Social perceptions and attitudes of children." *Genet. Psych. Monogr.* 40, 327–447, 1949.

INSIGHT INTO ONE'S OWN VALUES*

JULIAN C. STANLEY

George Peabody College for Teachers

Twenty years have elapsed since Vernon and Allport[17] first published an attitude scale, *A Study of Values*, designed to validate Eduard Spranger's[14] six 'types of men' empirically. Except for its social-value scores, which are quite unreliable, the instrument has proved to be useful in many fields, particularly for comparing groups. A considerable amount of normative data showing characteristic patterns for various occupations and institutions has appeared in the psychological literature.

SELF-RATINGS

One aspect of validity, correlation of self-rankings with scores on the scale, has received comparatively little attention. In their original article Vernon and Allport[17] reported *r*'s between scores and "the average of five external and one self-rating" for each person in "a fairly heterogeneous group" of forty-eight college students (pp. 245, 242). These are shown in Table 1.

Two years later Pintner[12] obtained correlations with self-ratings for one hundred eighty-seven students in an educational psychology class (fifty-eight men, one hundred twenty-nine women), which also appear in Table 1. He concluded, "With the exception of political interests, all these values are positive and some are very satisfactory considering the unreliability of the self-ratings" (p. 356).

Recently Anderson[2] studied self-rankings of two hundred eight male vocational counseling clients "prior to the interpretation of their results on the Allport-Vernon *Study of Values*" and after these individuals "were given an explanation of the six directions of interest represented in this measure" (p. 350). Subjects did not rank all six values, but instead indicated "the two among the six values which they judged to be most significant to them" and "the two among the six values which they judged to be

---

* The writer is indebted to Doctors Gordon W. Allport, Susan W. Gray, and John W. Gustad for suggestions that proved helpful to him in reporting the results of this study.

least significant to them" (p. 350). Her analysis of this data is couched in terms of percentages rather than correlation coefficients, so no direct comparison with the above studies is possible. The general trend seems to be essentially the same, however. Of particular interest is her finding that self-understanding apparently increases with age, the 26+ group exceeding the 18–25 clients.

TABLE 1.—CORRELATION BETWEEN SELF-RATINGS AND STUDY OF VALUES SCORES IN THREE INVESTIGATIONS

Study	Value					
	Theoretical	Economic	Aesthetic	Social	Political	Religious
Pintner[12]	.31	.39	.59	.14	−.02	.68
Vernon and Allport[17]	.40	.57	.57	−.06	.44	.69
This Study	.31	.51	.51	.15	.24	.51

Although the SV* is a forced-choice instrument which indicates the strength of the six evaluative attitudes relative to each other, apparently the rating scales used by Pintner[12] and Vernon and Allport[17] did not employ a ranking technique, but instead allowed the subject to rate himself high (or low) on every value if he so desired. Self-ranking without ties seems to the present writer more in harmony with the rationale of the SV than are noncomplementary ratings. Also, untied ranks make feasible the computation of Pearson product-moment r's between scores and self-ranks for each individual, because certain constants can be inserted in the usual formula for the correlation coefficient with ensuing simplifications. This technique was employed in the present study.

SUBJECTS AND METHOD

The subjects were sixty-six male freshmen taking an elective course in general psychology at a small New England municipal junior college. They comprised about fifty per cent of all male freshmen at this institution and were reasonably representative of that group. The age range was 17–25 years, with a mean of 20.0 and a standard deviation of 1.8.

---

* Throughout the rest of this paper 'SV' will refer to the Allport-Vernon *Study of Values.*

On a Monday during the second semester each subject was given an SV booklet and allowed to complete it in class. The scale was not discussed until the next Monday when the writer suggested that self-ratings be made and when he read to the group the outline of Spranger's types from pp. 8–11 of the Manual of Directions[1] without additional comments. Every person was then asked to rank himself without ties from 1 (highest value) to 6 (lowest value) on the categories. Only after these rankings were collected did students learn their scores.

Both scores and rankings were secured for sixty-six of the sixty-eight men enrolled; two were absent when students ranked their values. Complete data were obtained for all twelve female members of the group, but because of the small number involved these results have not been included in the present report.*

RESULTS

Scores and Self-Rankings.—Table 2 summarizes SV scores for these sixty-six male freshmen, who were highest on political and economic values and lowest on aesthetic. Their profile is fairly similar to the typical one for males reported in the Manual of Directions.[1]

TABLE 2.—MEANS AND STANDARD DEVIATIONS OF STUDY OF VALUES SCORES FOR 66 MALE COLLEGE FRESHMEN

	Value					
	T	E	A	S	P	R
Mean	29.3	32.0	23.7	30.3	34.0	30.8
SD	6.1	6.8	7.6	5.0	7.0	7.2

Information concerning self-rankings is given in Table 3, where the highest average rank, 2.6, is on the social category and the lowest, 4.6, is on the religious. Correlating score means with rank means produces an r of only .15,* so for the group as a whole there was poor agreement. If, however, we compute the correlation between scores and ranks separately for each value, the r's are somewhat higher (see Table 1). These figures agree fairly well with those of Pintner[12] and Vernon and Allport,[17] even though Pintner (and perhaps Vernon and Allport, also) employed both sexes and should therefore have had more score variability and a consequent tendency toward larger r's than in the present investigation.

---

* Their self-ranking correlations did not, however, differ markedly from those for the men.

TABLE 3.—MEANS AND STANDARD DEVIATIONS OF SELF-RANKINGS ON STUDY OF VALUES CATEGORIES FOR 66 MALE COLLEGE FRESHMEN
(1 = highest rank, 6 = lowest)

	Value					
	T	E	A	S	P	R
Mean	3.2	2.7	3.9	2.6	3.9	4.6
SD	1.5	1.7	1.7	1.4	1.7	1.3

Individual Correlations.—Sixty-six correlation coefficients were computed, showing for each subject the relationship between his scores and ranks, and the median value was found to be .39, with a range from .98 to −.81. Since the distribution of these individual r's is negatively skewed, it was thought desirable to employ Fisher's[4] z-transformation with the correction for systematic positive bias (p. 205). In Table 4 will be found a distribution of the corrected z's as recommended for graphic portrayal by Kelley.[8] The mean z computed from ungrouped data is .369, to which corresponds an r of .35. The difference between this mean z and a z of 0 is more than five times its standard error. The sample standard deviation, .54, approximates the theoretical standard error of a z based upon six pairs, $1/\sqrt{3} = .58$.

---

* Because the highest rank is 1 and the lowest 6, negative coefficients of correlation between ranks and SV scores have a positive meaning in a 'high ranking goes with high score' sense (and, conversely, r's that are initially positive have a negative connotation). Therefore, in this study all r's between ranks and scores have been multiplied by −1 before being reported.

Intra-individual variability, represented by the standard deviation of the subject's SV scores, correlates .38 with the z's, showing in this study a significant though slight tendency for persons with the more pronounced value patterns to rate themselves better than do those with less dispersion. Inspection of the scatter diagram for this r reveals that the regression is probably rectilinear.

Reliability.—Vernon and Allport[17] listed split-half reliability coefficients based upon their 'fairly heterogeneous group' of forty-eight college students but did not indicate the methods of splitting employed. Careful scrutiny of the SV score sheet shows that a split producing forms comparable with respect to content is very difficult, if not impossible, to secure. A simple odd-even division of the present data resulted in the following critical ratios for the differences between means of the halves: theoretical, 0.3; economic, 7.2; aesthetic, 4.6; social, 3.7; political, 2.1; and religious, 4.1. All but one of these CR's are significant beyond the five per cent level, so the halves do not fulfill the assumption of equal means set forth by Jackson and Ferguson[7] (p. 110). Nevertheless, for comparative purposes r_{II}'s determined from these halves* are shown with the Fisher and Vernon-Allport

TABLE 4.—FREQUENCY DISTRIBUTION OF 66 CORRECTED z'S ($i = .41$)

Class Limits	Class Index	r Equivalent of Class Index	Frequency
1.85 & 2.25	2.05	.97	1
1.44 & 1.84	1.64	.93	1
1.03 & 1.43	1.23	.84	0
.62 & 1.02	.82	.68	23
.21 & .61	.41	.39	15
− .20 & .20	0	0	16
− .61 & − .21	− .41	− .39	8
−1.02 & − .62	− .82	− .68	1
−1.43 & −1.03	−1.23	− .84	1

---

* Instead of computing the r between scores on the two halves and 'stepping this up' by means of the Spearman-Brown formula, the writer employed Rulon's[13] simplified procedure.

coefficients in Table 5, where it will be seen that with the exception of the political value all of the present figures are lower than those reported in 1931. This may perhaps be attributable in part to the heterogeneity, especially with respect to sex, of the earlier group.

TABLE 5.—STEPPED-UP SPLIT-HALF RELIABILITY COEFFICIENTS FOR THE *Study of Values* OBTAINED IN THREE DIFFERENT INVESTIGATIONS

	Value					
	T	E	A	S	P	R
Fisher:[5]						
Fathers	.50	.71	.64	.45	.43	.75
Mothers	.65	.60	.39	.57	.43	.75
Sons	.49	.72	.81	.59	.58	.74
Daughters	.69	.64	.62	.58	.37	.76
Vernon and Allport[17]	.62	.72	.84	.49	.53	.84
This Study	.45	.64	.78	.01	.72	.46

Harris,[6] Todd,[16] and Whitely[18] report substantial test-retest *r*'s even for the social value, so the low split-half coefficients found in the present study can probably be charged in some measure to the method of splitting and to error variance in the academically mediocre group of freshmen taking the test. The entire SV scale is quite short, however, so reliability coefficients for each of the six categories cannot reasonably be expected to be high enough for individual predictive purposes.

DISCUSSION

Vernon and Allport[17] remark that "Empirical validity cannot be found satisfactorily by comparing the test results with ratings, owing to the unfamiliarity of the average rater with the conceptual nature of the values. His misinterpretation of the social value, combined with the ambiguity inherent in Spranger's conception of this value, is fatal" (p. 245). Pintner's[12] subjects misinterpreted the political category (direct expression of power) even more than the social.

Most available evidence seems to indicate that the social category lacks reliability. The political value is reasonably reliable but sometimes exhibits quite low correlations with ratings. In a personal communication to the writer, Dr. Allport remarked concerning these observations and others: "If the SV has high validity (and it seems to have) but if self-ratings are relatively poor against this criterion, does it not follow that students require an opportunity to get better insight? In other words, the SV is a good didactic device. What students need is to compare themselves with their contemporaries on a common scale. Until they do so their self-knowledge is deficient." With this point of view the present writer concurs fully.

The z's probably have rather low reliability, but this conjecture has not been tested in the present study. Such a determination would involve two administrations of the SV to each subject, followed by two sets of self-rankings for each person, and would also permit computation of an r_{11} for the ranks.

If z is an estimate of the subject's self-insight into his value-system as defined by Spranger, what characteristics should persons with high z's have that differentiate them from those with low z's? We might surmise that the former individuals would react more objectively to their environments, rationalize less, and in general be better adjusted. They should not necessarily be superior students, however, since the insightful individual whose highest value is, say, economic, may be content barely to pass his work in order to meet minimum standards for a degree that will have monetary worth for him. In fact, none of the values, with the possible exception of the theoretical one, which resembles Thurstone's[15] 'interest in science,' seems logically related to desire for general academic success.

Since good reading ability appears to be necessary in order to fill out the SV intelligently, and since understanding the description of Spranger's types as set forth in the Manual of Directions calls for considerable knowledge of vocabulary, the possibility that the z's are positively associated with scores on a reading test was investigated. The r between z's and total scores on the Nelson-Denny Reading Test,[10] which is speeded and contains both vocabulary and paragraph-comprehension subtests, was only .10, while the z's correlated .14 with scores on the Revised

Minnesota Paper Form Board Test,[9] an essentially nonverbal measure of spatial-relationship ability.* Neither of these r's differs significantly from zero, of course.

Vernon and Allport[17] point out that intra-individual variability is one measure of the reliability of profiles on the six values. For poor readers the scale probably should be less reliable than for those who read better, so we might expect moderate positive correlation between individual σ's and scores on the Nelson-Denny Reading Test, but this r is just .10. Evidently the range of reading ability among these subjects is not great enough to affect intra-individual variability very much.

Future Research.—Both z's and individual σ's seem worthy of further study, especially with reference to personality variables. It would be helpful to know, for instance, how self-insight as measured by z varies from the freshman to the senior year of college and what change occurs in σ during that interval. A carefully planned longitudinal investigation would be needed to follow up Anderson's[2] observation that self-insight appears to increase markedly with age. On *a priori* grounds possibly we should hypothesize that both z and σ are thus related to chronological age. It is not sufficient to establish that in a given group significant correlation exists between z's and age, because the older persons may differ considerably from the younger ones in many respects other than age.† For example, we cannot safely assume that when fifty-year-olds in such a group were twenty their z's would have been of the same magnitude as those of the current twenty-year-old subjects.

Among students in the present study, as in most such school-grade situations, age is negatively correlated with tested intelligence. If both chronological and mental age are positively correlated with z, some eighteen-year-olds may be able to compensate for youth with brightness and therefore earn higher z's than older classmates. Furthermore, practically all of the older students are World War II veterans, while few of the younger ones have had such military experience. These factors point toward the necessity for following the same persons through college and, preferably, into later life.‡

---

* The r between NDRT and RMPFBT scores for this group is .35.

† For the relatively homogeneous group utilized here, the r between chronological age to the nearest year and z is .14.

‡ For a thorough discussion of this general point see Corey.[3]

Statistical Note.—Peters and Van Voorhis[11] call the r between scores and ranks *rho prime* and state that Pearson's correction formula for estimating the r that would have been obtained if scores instead of ranks had been available is $r = 1.0233\rho'$ (p. 109). Since this adjustment is small and depends upon an untested assumption, it was not made in the present study.

The mean of six untied ranks is 3.500 and the standard deviation 1.708. The mean of an individual's SV scores is always 30.000, so the customary formula for the product-moment coefficient of correlation may be simplified for our purposes to

$$\frac{630 - \Sigma XY}{4.18\sqrt{\Sigma Y^2 - 5400}},$$

where X designates any rank and Y any SV score. It will be noted that because a rank of 1 is 'high' and 6 'low,' we have multiplied the usual formula by -1 in order that a tendency for high ranks to go with high SV scores will result in a positive coefficient.

SUMMARY

After reviewing the brief literature dealing with self-rating on the Allport-Vernon *Study of Values*, the writer discusses a study involving self-rankings without ties made by sixty-six male college freshmen. For individuals the average correlation between these ranks and SV scores was approximately .35. In the group as a whole social-value scores correlated lowest with self-rankings (.15), while aesthetic, economic, and religious values were highest and tied with .51.

In this investigation there is a slight tendency ($r = .38$) for students whose values are pronounced to have better 'self-insight' than those whose profiles are relatively flat.

A measure of insight called z, which can be computed fairly quickly with the assistance of an electric calculator, is offered for further study, especially with regard to its reliability and validity.

Some measure of intra-individual variability such as the standard deviation is suggested for inclusion in future research projects involving the *Study of Values*.*

---

* Since this paper was written, Allport and Lindsey have revised and restandardized the SV, giving particular attention to the hitherto unsatisfactory social scale.

BIBLIOGRAPHY

1) Gordon W. Allport and Philip E. Vernon. *A study of values.* Booklet, score sheet, and manual of directions. Boston: Houghton Mifflin Co., 1931.

2) Rose G. Anderson. "Subjective ranking versus score ranking of interest values." *Personnel Psychol.*, 1948, 1, 349–355.

3) Stephen M. Corey. "Attitude differences between college classes: a summary and criticism." *J. educ. Psychol.*, 1936, 27, 321–330.

4) Ronald A. Fisher. *Statistical methods for research workers.* (10th ed.) London: Oliver & Boyd, 1948.

5) Sarah C. Fisher. "Relationships in attitudes, opinions, and values among family members." *Univ. Calif. Publ. in Culture & Society*, 1948, 2, 29–99.

6) Daniel H. Harris. "Group differences in values within a university." *J. abnorm. soc. Psychol.*, 1934, 29, 95–102.

7) Robert W. B. Jackson and George A. Ferguson. "Studies on the reliability of tests." *Bull. No. 12, Dept. Educ. Res., Univ. of Toronto,* Toronto 5, Ontario, 1941.

8) Truman L. Kelley. *Fundamentals of statistics.* Cambridge, Mass.: Harvard Univ. Press, 1947.

9) *Minnesota Paper Form Board Test, The Revised.* New York: Psychological Corp.

10) *Nelson-Denny Reading Test, The.* Boston: Houghton Mifflin Co.

11) Charles C. Peters and Walter R. Van Voorhis. *Statistical procedures and their mathematical bases.* New York: McGraw-Hill, 1940.

12) Rudolph Pintner. "A comparison of interests, abilities, and attitudes." *J. abnorm. soc. Psychol.*, 1933, 27, 351–357.

13) Phillip J. Rulon. "A simplified procedure for determining the reliability of a test by split-halves." *Harv. educ. Rev.*, 1939, 9, 99–103.

14) Eduard Spranger. *Types of men.* (Trans. by P. J. W. Pigors.) Halle (Saale): Max Niermeyer Verlag, 1928.

15) Louis L. Thurstone. "A multiple factor study of vocational interests." *Personnel J.*, 1931, 10, 198–205.

16) J. Edward Todd. *Social norms and the behavior of college students.* Contributions to Educ., No. 833. New York: Bureau of Publications, Teachers College, Columbia University, 1941.

17) Philip E. Vernon and Gordon W. Allport. "A test for personal values." *J. abnorm. soc. Psychol.*, 1931, 26, 231–248.

18) Paul L. Whitely. "A study of the Allport-Vernon test for personal values." *J. abnorm. soc. Psychol.*, 1933, 28, 6–13.

METHODS FOR INDICATING THE TRUE SIZES OF PICTURED OBJECTS[1]

MORTON S. MALTER

Michigan State College

Textbook illustrations may depict objects in enlarged, actual or reduced sizes. Occasionally, related objects similar in size are reduced by the same proportions and illustrated in one panel. Readers familiar with one object can then use it to judge the sizes of the others. More often no attempt is made to reduce objects by the same ratios because they are too diverse in size. For example, an elephant might be reduced in a ratio that would result in an adequate sized illustration. To reduce other objects similarly would result in exceedingly small representations. Likewise, the attempt seldom is made to enlarge items by the same proportion. Furthermore, within many books some objects necessarily are enlarged or reduced, while others lend themselves to reproduction in actual size. The result is that readers seldom have an opportunity to utilize familiar pictures in evaluating the actual sizes of unknown representations.

To guide readers in judging the actual sizes of representations, captions are used. These captions sometimes indicate actual size in ratio form, e.g. "¼ actual size." Ratios are unacceptable to most educators for they generally are difficult to interpret. More often the captions are in terms of linear measurement, e.g. "The object is 5 feet long." Although this type of caption is assumed to be useful for adults, some educators have questioned its validity for younger pupils. They would prefer, instead, that the size of a pictured object be shown through comparing it to a familiar item. Dale,[2] for example, states: "Unless the picture includes one item whose dimensions are familiar to students, it will be impossible for them to know the sizes of the objects. A good photograph of the pyramids includes a man or a camel, and immediately the child appreciates the relative sizes. It is not enough to tell pupils how large the items are. . . . "

---

[1] This study was subsidized in part by a grant from *Britannica Jr.*

[2] Edgar Dale, *Audio-Visual Methods in Teaching*, New York: Dryden Press, 1947, p. 230.

This writer could find no studies concerned with the ability of pupils to utilize either linear measurement or familiar objects in judging the sizes of unknown representations. This study, therefore, was conducted to determine the relative effectiveness of these two methods. The study is limited to pupils in grade VI. It involves only illustrations whose referents range between six inches and six feet. The illustrations, furthermore, are limited to animals reduced by varying proportions.

PROCEDURE

For purposes of the experiment, the following null-hypothesis was formulated: Numerical terms (linear measurement) are not significantly more useful than familiar objects for helping pupils in grade VI judge the actual sizes of unknown representations. The hypothesis was tested as follows:

Table I.—The Fourteen Animals Represented in Pictures

Picture Number	Name	Geographic Habitat	True Size
1	Adjutant Bird	India, Africa	5 feet
2	Magpie	Western U.S.A.	20 inches
3	Tapir	Asia	3½ feet
4	Wart Hog	Africa	27½ inches
5	Flamingo	S. America Cen. America	4 feet
6	Flycatcher	U.S.A.	14 inches
7	Yak	Asia	6 feet
8	Horned Toad	Western U.S.A. Mexico	6 inches
9	Flying Fish	Tropical Waters	18 inches
10	Aardvark	Africa	5 feet
11	Marten	U.S.A., Europe	18 inches
12	Civet Cat	Asia, Africa	2½ feet
13	Barracuda	Tropical Waters	3–8 feet
14	Gila Monster	Western U.S.A.	2 feet

Fourteen pictures of animals were selected from a juvenile encyclopedia. (Table I.) The pictures were submitted to a group of grade VI pupils who were asked to identify each picture. The results indicated the pupils were unfamiliar with each of the fourteen animals. Pupils tested in this preliminary survey were selected from populations similar to those the writer planned to use in the experiment. It was assumed, therefore, that the fourteen animals represented unfamiliar animals. These fourteen drawings were designated as Set No-Help.

A pair of silhouettes was inserted in the right-hand corner of a copy of each drawing in the Set No-Help. One member of each pair represented the central animal in the drawing proper. The other member represented a common object, i.e., a man, an arm or a fork, drawn to scale and intended to indicate relative size. It was assumed that the silhouette device would simulate a drawing proper including a familiar object. In fact, the silhouette device was probably more advantageous, since the possibility of perspective confusing the reader was minimized through placing both objects in the pair on the same plane.

The silhouette devices in pictures 1–9, 11 and 13–14 were intended to be of maximum value to the pupils. In these twelve sets the familiar objects' most familiar dimension, e.g., the height of man, was inserted parallel with the dimensions of the animals the pupils were expected to judge. Furthermore, in these sets the sizes of the familiar and unknown objects were not divergent. In pictures 10 and 12 the familiar objects were represented vertically, although pupils were expected to estimate the horizontal dimensions of the animals. In picture 12 the familiar object was much larger than the animal. These fourteen drawings were designated as Set Silhouette-Help.

Copies of each drawing in Set No-Help were used to prepare a third exhibit. This time the size of each animal was indicated by a linear measurement caption inserted beneath each picture. The captions were taken exactly as stated from the juvenile encyclopedia containing the original drawings. The last column in Table I shows the numerical caption included with each picture. These fourteen drawings were designated as Set Numerical-Help.

Sufficient copies of each of the three sets were planographed to provide proper materials at the desk of each member of a class being tested. All factors in the drawings proper were kept constant in the planographing process.

A desk chart was next constructed as follows: A line measuring ten and one-half inches in length was placed on paper. The line was marked off and numbered in eighty-seven equally spaced places. Each number was inserted twice to enable one to read the numbers when the chart was held either vertically or horizontally. The desk chart was then planographed.

A vertical wall chart, measuring eleven inches across and one hundred four inches in height was designed. The chart was marked off every four inches in numbers from 101 to 127. Number 101 appeared four inches from the bottom of the chart and the remaining numbers followed consecutively upwards. A horizontal wall chart of the same dimensions was likewise constructed. This chart was marked off every four inches in letters from A at the left to Z at the right. The letter A was inserted four inches from the left-hand margin. The markings on both charts were sufficiently legible to be read from all parts of a typical classroom.

A multiple-choice test was constructed on the basis of a pilot study conducted in three Grade VI classes. Pupils in the pilot study responded to the fourteen drawings in the same manner described later in this study. Their free responses were used to develop the items for the multiple-choice test. The item developed in connection with each picture is shown in the column marked 'Choice' in Table II. The following section describes how the multiple choice test was used along with the drawings, desk chart and wall charts in collecting the data.

COLLECTING THE DATA

The method of administering the materials is considered for Class 1, a Grade VI class in Illinois. On a designated date, the investigator took charge of the class, consisting of twenty-four pupils. The vertical and horizontal wall charts were suspended in front of the room. The pupils were divided at random into two groups. All pupils were given a desk chart, multiple choice test, and the fourteen drawings in Set No-Help. They were given the following directions:

> Look at Picture 1. It is a picture of a bird. Notice that the bird is standing on the ground and that the top of its head is marked off in dash lines. I want you to imagine how tall the bird is in real life, between the part touching the ground and the part marked off in dash lines. Next, look at the wall chart marked off in numbers. I want you to imagine that the bird, in real life, is standing against the wall

TABLE II.—NUMBER OF PUPILS IN EACH GROUP SELECTING CHOICES FOR EACH PICTURE

Picture	Choice	Inches	Num.H.	Sil.H.
1	a) 103	12	2	0
	b) 106	24	2	1
	c) 109	36	9	4
	d) 112	48	10	7
	*e) 115	60	40	34
	f) 115+	60+	13	30
2	a) −34	−4	14	5
	b) 34	4	20	5
	c) 103	12	16	11
	*d) 105	20	21	31
	e) 107	28	3	17
	f) 109	36	2	7
3	a) 103	12	7	0
	*b) 109	36	57	69
	c) 115	60	11	5
	d) 118	72	1	1
	e) 121	84	0	1
	f) 121+	84+	0	0
4	a) 105	20	20	7
	*b) 107	28	25	24
	c) 109	36	14	33
	d) 111	44	13	12
	e) 113	52	3	0
	f) 113+	52+	1	0
5	a) −109	−36	4	0
	b) 109	36	10	1
	*c) 112	48	38	27
	d) 115	60	13	37
	e) 118	72	7	10
	f) 121	84	4	1
6	a) −103	−12	21	12
	*b) 103	12	26	34
	c) 105	20	22	18
	d) 107	28	5	5
	e) 109	36	1	1
	f) 111	44	1	6
7	a) 106	24	1	1
	b) 109	36	5	1
	c) 112	48	13	14
	d) 115	60	12	20
	*e) 118	72	35	32
	f) 118+	72+	10	8
8	a) 18	2	16	10
	*b) 51	6	35	42
	c) C	12	18	17
	d) D	16	6	2
	e) E	20	0	3
	f) E+	20+	1	2
9	a) −68	−8	6	6
	b) 68	8	7	7
	*c) D	16	23	12
	d) F	24	22	25
	e) H	32	8	19
	f) J	40	10	7
10	a) C	12	0	5
	b) F	24	1	6
	c) I	36	9	23
	*d) O	60	36	21
	e) U	84	26	18
	f) U+	84+	4	3
11	a) −51	−6	4	7
	b) 51	6	4	2
	*c) D	16	29	11
	d) G	28	20	35
	e) I	36	15	18
	f) L	48	4	3
12	a) −51	−6	4	7
	b) 51	6	6	14
	c) C	12	6	20
	*d) F	24	47	29
	e) L	48	11	5
	f) O	60	2	1
13	a) 51	6	3	5
	b) C	12	0	1
	c) F	24	11	4
	*d) O	60	39	38
	*e) X	96	18	34
	f) X+	96+	5	4
14	a) −26	−3	1	1
	b) 26	3	5	8
	c) 51	6	9	6
	d) C	12	9	10
	*e) F	24	42	38
	f) L	48	10	13

* Correct Answers

chart with its feet on the ground. Now look at the multiple-choice test. Notice that there are six choices under the part marked Picture 1. The choices are marked a, b, c, d, e, and f. Here is what the choices mean. Choice "a" means the bird shown in Picture 1 would reach to number 103 if it were standing against the wall chart. Choice "b" means that the. . . .

Similar directions were given for Picture 2, which involved using the desk chart and vertical wall chart. After explaining the directions for Pictures 1 and 2, the investigator found the pupils were able to proceed on their own in judging the sizes of the remaining animals. A number of pupils requested individual attention at their desks. Among the questions raised were: 1) "Between which two parts of the animals do I estimate real life size?" 2) "How do you use the multiple-choice test?" 3) "Do we imagine the animal is standing on the floor or on No. 101?" The writer is convinced that his answers clarified the questions.

When all subjects had completed the test, the drawings and multiple-choice sheets were collected. Blank multiple-choice sheets were given again to all pupils. Twelve subjects selected at random were given Set Silhouette-Help; the remaining subjects were given Set Numerical-Help. All pupils were asked again to estimate the sizes of the animals, utilizing the help provided with the drawings. The subjects were aware only of the help provided with their particular set, since the groups previously had been placed in different parts of the room. The pupils experienced no difficulty in following the directions.

The fact that subjects in the groups were seated at varying distances from the wall chart was discounted as unimportant to the experiment on the basis of two assumptions. First, the assumption was made that equal numbers of pupils in both groups would be seated at the same distance from the wall chart. Secondly, the factor of size constancy was assumed. The term 'size constancy' is used to refer to the fact that the seen sizes of object do not diminish as they become more distant, as do the retinal images corresponding to these objects. In this instance, the 'seen sizes' of objects refers to distances on the wall chart, against which the subjects had projected mental images of the animals.

Data were collected in a like manner from five other Grade VI classes, making a total of one hundred fifty-four pupils tested in the experiment. The IQ of the seventy-six pupils provided with Set Silhouette-Help was 103.26 (σ 11.17); their CA was 139.79 months (σ 7.34). The IQ of the seventy-six pupils provided with Set Numerical-Help was 103.73 (σ 11.09); their CA was 140.56 months (σ 9.12).

TEST RESULTS

To test the formulated hypothesis, only the second multiple-choice papers submitted by all pupils were utilized.[3] Data on these papers were analyzed in three ways. First, each paper in both groups was marked according to number correct. There was only one correct choice allowed for each test item with the exception of number 13 (Table II). This animal ranged between sixty and ninety inches according to the caption, and both choices 'd' and 'e' were considered correct. The lowest and highest score that may have been assigned to any paper were 0 and 14. The mean scores for the pupils using Set Numerical-Help and Set Silhouette-Help were 7.18 and 6.88, respectively. The significance of the difference between the means of the two small samples was computed and $t = 1.15$. For ∞ *d.f.* a t of this value may have occurred by chance variation between twenty and thirty per cent of the time. Thus, the hypothesis that the numerical terms are not significantly more useful than familiar items in helping pupils in Grade VI judge the true sizes of unknown representations was accepted with a reasonable amount of confidence.

In the t test just described only one correct choice was assigned to each individual item. The writer realized, however, that more than one choice could be scored as reasonably correct. For example, the Adjutant Bird in Picture 1 is designated as five feet tall, but this figure is only the mean of a normal distribution. Choices 'd' (three feet), 'e' (five feet) and 'f' (five feet plus) are reasonable, and pupils arbitrarily could be given credit for marking them. As a second means of testing the formulated null-hypothesis, the percentages of pupils in each group reasonably estimating the true size of each pictured object were compared.

---

[3] The first multiple-choice paper was used to determine whether the numerical terms and silhouette devices helped pupils judge the actual sizes of pictured objects. The results showed both methods significantly helped the pupils. See Morton S. Malter, *The Relationship of Certain Variables to Children's Ability to Estimate the True Size of Pictured Objects*, (Unpublished), University of Chicago, Chicago, Ill.

In making this comparison, a formula developed by McNemar[4] for calculating the statistical significance of the difference of two non-correlated percentages were used. Table III designates whether or not the differences between the various pairs of percentages were significant at the one per cent level of confidence. Thus the silhouette device in Picture 3 seems to be more useful than the corresponding caption (5 feet). The numerical terms under Pictures 10 and 12 seem to be more useful than the corresponding silhouette devices. This finding is expected in view of the manner in which the silhouette devices in Pictures 10 and 12 were constructed. None of the differ-

TABLE III.—PERCENTAGES OF PUPILS IN GROUPS REASONABLY ESTIMATING THE TRUE SIZES OF THE DESIGNATED PICTURES

Picture	Reasonable Estimates	Numerical Help	Silhouette Help	.01 Probability
1	def	83	93	—
2	d	28	41	—
3	b	75	91	x
4	abc	77	84	—
5	bcd	80	86	—
6	ab	62	61	—
7	def	75	79	—
8	b	46	55	—
9	cd	59	49	—
10	d	47	28	x
11	cd	64	60	—
12	d	62	38	x
13	def	82	87	—
14	de	67	63	—

---

[4] Quinn McNemar. "Note on the Sampling Error of the Difference between Correlated Proportions or Percentages," *Psychometrika*, XII (1947), 153–57.

ences between the other pairs of percentages were statistically significant. One may conclude that, in general, the Numerical-Help and Silhouette-Help methods are not significantly different with respect to providing pupils in Grade VI with a reasonable idea of the true size of a pictured object.

The data so far have been analyzed from two standpoints. On the basis of the results it was concluded that the two methods did not differ significantly with respect to communicating either a correct or reasonable idea of the true size of a pictured object. The data were treated in a third manner to determine whether or not the methods varied significantly with respect to affecting all choices pupils made on the multiple-choice test. All choices are now under consideration, whether or not they are correct or reasonable.

The chi-square was computed for each contingency table distribution listed in Table II. Table IV lists the picture distributions according to numbers and the corresponding probabilities. The chi-square for distribution 1 is 12.94. For 3 *d.f.* a chi-square of this value would be exceeded much less than .005 of the time. An aggregate chi-square computed for the test in general was found to be 84.47. For 28 *d.f.* a chi-square of this value (or larger) would be exceeded much less than .01 of the time. One may conclude, therefore, that the two methods do differ significantly with respect to affecting the estimates pupils make of the size of pictured objects.

This latter conclusion must be interpreted cautiously. By no means does the conclusion imply either of the two methods is better. Nor does it imply that the entries in the adjoining correct cells are significantly different. It does indicate that entries in certain of the adjoining cells in the various contingency tables did vary significantly. Table IV suggests that the distributions demanding our inspection are 1, 2, 4, 5, 10 and 12.

Examination of distributions 1, 2, 4 and 5 (Table II) suggests the pupils in the Numerical-Help Group and Silhouette-Help Group were slightly underestimating and overestimating respectively. Part of this finding is corroborated elsewhere by the writer,[5] who found that pupils in Grade VI tend to underestimate numerical terms of considerable length, e.g., more than three feet. The fact that the distributions for Pictures 10 and 12 were significantly different again was expected. The silhouette devices in these two pictures were purposely constructed in a manner considered perverse, and thus the entries in corresponding cells in the Numerical Help and Silhouette columns were extremely divergent.

---

[5] Malter, *op. cit.*, p. 112.

TABLE IV.—CHI-SQUARES AND CORRESPONDING PROBABILITIES ON PICTURE ITEM DISTRIBUTIONS OF THE NUMERICAL HELP AND SILHOUETTE HELP GROUPS

Picture	Category	Chi-Square	*d.f.*	Probability
1	(abc)-d-e-f	12.94	3	.005
2	(ab)-c-d-(ef)	28.38	3	.004
3	(ab)-(cdef)	1.50	1	.200
4	a-b-c-(def)	11.68	3	.009
5	(abc)-d-(ef)	18.72	2	.004
6	a-b-c-(def)	5.24	3	.157
7	(abc)-d-e-f	2.62	3	.460
8	a-b-c-(def)	2.04	3	.565
9	(ab)-c-d-e-f	8.66	4	.715
10	(abc)-d-(ef)	18.60	2	.001
11	(ab)-c-d-(ef)	12.34	3	.064
12	(abc)-d-(ef)	18.28	3	.004
13	(abc)-d-(ef)	1.64	2	.452
14	(ab)-c-d-e-f	1.86	4	.757

CONCLUSIONS AND IMPLICATIONS

The results indicate there is no significant difference in the effectiveness of selected familiar objects and numerical terms for communicating to pupils in Grade VI the actual size of unfamiliar representations. Pupils in this grade are sufficiently acquainted with the meaning of linear measurements ranging from six inches to six feet to use them effectively. The pupils, of course, varied in ability to assign a correct interpretation to a specified measurement. However, they also varied in ability to interpret correctly designated familiar objects. An implication is that publishers of Grade VI materials need not insert familiar objects in pictures to indicate the sizes of unknown items ranging between six inches and six feet. Numerical terms suffice for this purpose.

The results possibly may have been altered in favor of the silhouette-devices if different familiar objects were used, especially in Pictures 10 and 12. The writer believes, however, that the familiar objects selected were reasonable. In fact, the experiment possibly was biased in favor of the silhouette devices, since the possibility of perspective confusing the reader was minimized. Furthermore, the result may have been altered in favor of the numerical terms if different captions were used. For example, the caption under Picture 6, "14 inches," may have been more effective if stated as "1 foot, 2 inches." In general, the numerical captions and familiar items are representative of those which editors would select.

The writer realizes the plausibility in the argument that a picture suggesting relative size is acceptable at 'face validity.' In other terms, it may be sufficient for a pupil to realize an object stands in a certain relation to a man, regardless of his concept of the latter. However, it is equally reasonable to argue that association of an object with its correct linear measurement is also sufficient. A realization of his association has two advantages. First, the pupils can communicate or otherwise use the number with reference to the object. Secondly, the number may be translated into an exacting concept with the use of a standard yardstick.

No data were collected on the ability of pupils below Grade VI to use the alternatives. A concomitant study on the ability of pupils in Grade IV to estimate a series of distances between one inch and seven feet was conducted.[6] The results showed there was little differences between the mean estimates of these distances made by pupils in Grades IV and VI. There was, however, a significant difference in variability of estimates, with pupils in Grade VI more closely grouped about the mean. While this suggests numerical terms may be less effective in Grade IV, a study in this grade on the ability of pupils to estimate the size of familiar items may also indicate more variability.

The silhouette devices in Pictures 10 and 12 were intended to provide suggestions for future studies. In both sets the pupils were asked to estimate the horizontal dimensions of the animals, although the familiar dimension of the known object was shown vertically. The results suggest this method of presentation is undesirable. A more elaborate study to test this tenable hypothesis is in order.

---

[6] *Ibid.*, pp. 111–115.

Restrictions within the experiment prevented testing the formulated hypothesis in connection with objects of large magnitude. The writer believes that Dale is correct in contending that telling the size of a pyramid in terms of measurement is not likely to help the young reader. However, one must be cautious in assuming that a small familiar item, one most likely within the experiences of a reader, always suggests the actual sizes of a large object alongside it. Data collected in connection with Picture 12 suggests a familiar item may not be too effective when used to indicate the size of a much larger object. A study to determine the usefulness of small objects in indicating the size of large ones is necessary. In the meantime, the writer would suggest that both familiar items and captions be used to indicate the size of large objects. The advantage of the latter method, again, is that measurement can always be converted into meaningful concepts.

PREJUDICE VERBALIZATION AMONG CHILDREN

CONRAD CHYATTE

Catholic University of America

DOROTHY F. SCHAEFER

Teachers College, Columbia University

and

MARTIN SPIAGGIA

New York University

INTRODUCTION

Prejudice verbalization (which hereafter will be referred to as PV for the sake of brevity) is defined as the production of invidious comments concerning an individual, group, organization or concept. As an index of anti-minority attitudes, the presence or absence of such verbalization can by no means be considered synonymous with the presence or absence of prejudice itself. Among very young children, its presence may only represent a verbal behavior pattern copied from parents, their surrogates and childhood associates, rather than true emotionally negative attitude toward a minority. Among older children and adults, its absence can merely signify a reluctance to be identified as holding a socially reprehensible point of view.

The fact still remains, however, that PV is often accepted as *prima facie* evidence of negative attitude. In the case of the naïve individual, particularly the child, anti-Negro or anti-Semitic expression at the verbal level by others within the individual's social sphere will almost certainly be apperceived as prejudice *per se*. For the child, it appears beyond question that PV by others is of importance in the etiology of any prejudices he may develop. Secondarily, it is also suggested that PV among young children, learned at least in part from parents operating in the permissive atmosphere of the family, may actually reflect adult attitudes more accurately than the responses of the parents themselves in direct interview.

It was toward the identification of personal, social and economic factors associated with PV among children that this study was directed. Work by Frenkel-Brunswik,[3] Meltzer,[6] and

Hartley[4] indicate that children's attitudes are positively correlated with those of their adult associates. Previous research by Holmes[5] indicating that southern college students exhibit prejudice more frequently than do northern college students thus leads to the expectation that southern children will be more prone to PV than northern children.

Radke and Sutherland[7] reported a study of children's attitudes toward Negroes and Jews in a small mid-western town where there were no Negroes and only one childless Jewish family. They found that statements of dislike for both minority groups increased with age.

Adorno, et al,[2] working with adults, were able to find no simple relation between income and attitudes toward minority groups. However, they were able to discern a tendency for the offspring of fathers in professional occupations, particularly religious and government occupations, to exhibit less ethnocentric, or anti-minority, attitudes than did those of fathers with other types of occupations. They found no statistically significant sex differences.

Ackerman and Jahoda[1] suggest the importance of 'cultural climate' as a determinant of anti-minority attitudes, pointing to Nazi Germany as representing a cultural climate which encouraged anti-Semitism. In the United States, they say, the cultural climate is not consistent, support for anti-minority feeling varying from one cultural subgroup to another. They feel that anti-minority behavior can be expected to show considerable dependency on specific in-group pressures. Thus, while much effort has been made to explore the relationship between prejudice and emotional disorder, it is suggested that prejudice, particularly at the verbal level, need not necessarily be associated with personality aberrations but, instead, may be conformity behavior in a given cultural climate.

THE PROBLEM

The present study was designed to determine whether or not PV in children was significantly related to differences in sex, geographic location, father's occupation, national origin, age and school grade.

POPULATIONS

Three sub-populations were studied: seventy southern boys, seventy northern boys and seventy northern girls. The age distribution (range: four to fifteen years), which because of matching is the same for each sub-population, is presented as Table I. None of the subjects concerned were Negro or Jewish.

TABLE I.—DISTRIBUTION BY AGE FOR EACH SAMPLE

Age	N
4	1
5	6
6	5
7	10
8	5
9	8
10	15
11	7
12	7
13	3
14	2
15	1
Total	70

TABLE II.—DISTRIBUTION BY FATHER'S OCCUPATION FOR EACH SAMPLE

Occupation	Southern Boys	Northern Boys	Northern Girls
Farmer	49	0	0
Professional	0	6	6
Business	4	3	4
Clerical and Sales	0	7	10
Skilled	1	22	23
Semi-skilled	2	19	20
Unskilled	7	6	3
Miscellaneous and Unknown	7	7	4
Total	70	70	70

The Southern Sample.—The seventy southern boys were interviewed on a trip along U.S. Highway 301, between Fayetteville, North Carolina and Petersburg, Virginia, in June, 1950. They were predominantly children of farmers. A breakdown for this group by father's occupation is found in Table II. No data relative to school grade or national origin were obtained for this sample.

The Northern Sample.—The seventy northern boys and seventy northern girls were interviewed in the five boroughs of New York City between June and September, 1950. The breakdown for this group by father's occupation is given in Table II. Distributions by school grade (range: pre-school to tenth-grade) are presented as Table III. A description of this sample in terms of national origin is given as Table IV.

TABLE III.—DISTRIBUTION BY SCHOOL GRADE FOR NORTHERN SAMPLE

Grade	Boys	Girls
Pre-school	5	4
Kindergarten	2	3
First	6	7
Second	9	9
Third	7	6
Fourth	6	19
Fifth	14	6
Sixth	12	7
Seventh	4	1
Eighth	3	4
Ninth	2	3
Tenth	0	1
Total	70	70

PROCEDURES

Collection of Data.—Prejudice verbalization was elicited through the use of simple verbal stimuli. Each child was approached and an attempt to establish an informal, permissive relationship between him, or her, and the interviewer was made. It was often necessary for the interviewer to participate in the activities of the child—games of ball, Chinese checkers, climbing in the playground, etc. In those cases where a group of children was encountered, only one child from the group was interviewed, and this was done apart from the rest in order to avoid any immediate influence of the child's playmates on his, or her,

TABLE IV.—DISTRIBUTION BY NATIONAL ORIGIN FOR NORTHERN SAMPLE

National Origin	Males	Females	Total
"American"	0	2	2
Armenian	1	0	1
Chinese	0	1	1
Czechoslovakian	0	1	1
Danish	1	0	1
English	1	5	6
English-German	1	0	1
English-Irish	0	1	1
French	1	0	1
French-Canadian	0	1	1
French-Italian	1	0	1
German	13	5	18
Greek	0	2	2
Greek-Italian	0	1	1
Irish	15	18	33
Irish-German	1	2	3
Irish-Italian	1	1	2
Italian	14	9	23
Japanese	1	0	1
Lithuanian	2	3	5
Polish	5	6	11
Polish-German	1	0	1
Portuguese	2	0	2
Puerto Rican	3	4	7
Russian	2	3	5
Russian-German	1	0	1
Scottish-Italian	0	1	1
Spanish	1	3	4
Swedish-Irish	1	0	1
Ukrainian	1	1	2
Total	70	70	140

verbalizations. In each case, at suitable points during the interview, the questions,

> Do you know what Jews are? What do you think of them?
> Do you know what Negroes are? What do you think of them?

were asked, and the responses noted. In some cases, where the word 'Negroes' was not understood, the words 'colored people' were substituted. In all cases, the children were interviewed outside of the home and away from any institutional situation, but were instead contacted singly in the streets, parks, playgrounds, along the highway, etc.

The presence or absence of anti-minority verbalization was determined by examining the interview protocol which was written immediately after each conversation. Protocols were not written during interview in order that freedom of expression on the part of children might be encouraged.

Typical of the remarks judged to be PV were:

> "If any of them come to my block, I'll beat the hell out of them."
>
> "They have black faces and I don't like to play with them."
>
> "They got all the money . . . they're crooks . . . you can't trust them."

On the other hand, a comment similar to many frequently encountered, and which was judged to indicate absence of PV, was that made by a girl attending a parochial school: "God created all of us . . . made us all alike. We must love everyone . . . " In this case, as in a number of others, some of the child's spontaneous comments pointed to an awareness of the problem of prejudice. Other responses judged to be without PV included:

> "They're all right."
>
> "They're the same like anybody else."

During interview, when the interviewer was uncertain about whether or not PV was actually present (for example: "They're OK, but I don't play with them."), further probing was made until the attitudinal direction of the comments was clear.

Statistical Analysis.—In determining the statistical relationship of PV to discrete variables like geographic location, sex, father's occupation and national origin, the technique for determining the significance of a difference between two percentages was employed. For example, when northern boys were compared with southern boys on anti-Negro verbalization, the significance of the difference between percentage anti-Negro verbalization found in each of the two sub-populations was determined. In the case of the continuous variables, age and grade, the significance of a difference between the means of two samples was found. For example, in the northern sample, the mean age of children with anti-Jewish verbalization was compared with the mean age of those without anti-Jewish verbalization, and the significance of the difference between the two means was assessed.

In the cases of both the discrete and continuous variables, significance was estimated through the use of *t* ratios which were calculated and interpreted from Smith's Probability Table of *t*[8] in terms of the acceptance or rejection of the null hypothesis.

SUMMARY OF RESULTS

Differences in prejudice verbalization by geographic location, sex, father's occupation and national origin are summarized as Table V which shows that:

1) No statistically significant sex differences in PV were found in comparing northern girls with northern boys, although in each PV category, anti-Jewish, anti-Negro and the combined anti-Jewish—anti-Negro—or both, the boys showed a consistently higher percentage of PV.

2) The southern boys exhibited significantly more of each type of PV than did the northern boys. Since northern boys consistently showed more PV than northern girls, and, since there was an equal number of cases in each of the three sub-populations, *t* ratios for the northern girls versus the southern boys, which will necessarily reflect significant differences, were not computed.

3) In the northern sample (boys and girls combined), children of professional fathers showed significantly less PV of each type than did the children of non-professional fathers.

4) A group comprised of all children in the northern sample with English, Irish, German or Polish national origin was found to have a significantly higher proportion of each type of PV than did all of the other children in that sample.

Differences in PV for the northern sample by age and grade are presented as Table VI.

5) While the mean age and grade of northern children with anti-Jewish verbalization were lower than those found for northern youngsters free of this kind of PV, the differences

were not found to be statistically significant. However, the mean age and grade of children falling in the anti-Negro and the combined anti-Jewish—anti-Negro—or both categories, were significantly lower than those of children who did not express these types of PV.

TABLE V.—DIFFERENCES IN PREJUDICE VERBALIZATION BY SEX, GEOGRAPHIC LOCATION, FATHER'S OCCUPATION AND NATIONAL ORIGIN

Type of Verbalization:		Anti-Jewish	t Ratio	Anti-Negro	t Ratio	Anti-Jewish, Anti-Negro, or Both	t Ratio
By Sex and Geographic Location	Per cent of 70 Northern Girls	35.71	1.21	30.00	1.25	41.43	1.88
	Per cent of 70 Northern Boys	45.71	5.64*	40.00	7.63*	57.14	5.69*
	Per cent of 70 Southern Boys	89.47†		91.43		94.29	
Northern Sample: By Father's Occupation	Per cent of 12 Children with Professional Fathers	00.00	9.81*	08.33	3.26*	08.33	5.60*
	Per Cent of 118 Children with Non-professional Fathers	44.92		38.14		54.24	
Northern Sample: By National Origin	Per Cent of 68 Children with English, German, Irish or Polish National Origin	57.35	4.11*	45.59	2.60*	63.24	3.33*
	Per Cent of 72 Children with Other Types of National Origin	25.00		25.00		36.11	

* Significant at the one per cent level of confidence.

† This represents the percentage of southern boys who both knew the meaning of the word 'Jew' and exhibited anti-Jewish verbalization. Thirty-two Southern boys who did not know the word 'Jew' were excluded from the computation of this percentage.

In general (Table V), anti-Jewish verbalization tended to be more prevalent within the northern sample than was anti-Negro verbalization. In the southern sample, both types of PV were high and approximately equal percentage-wise. However, in no case was any sample significantly more anti-Jewish in PV than anti-Negro, *t* ratios concerned with this aspect hovering near zero. These *t* ratios are not recorded elsewhere in this report.

DISCUSSION

In the statistical analysis it was found that of all possible dichotomies based on father's occupation for the northern sample, the Professional—Non-professional dichotomy was the only one which yielded significant differences in PV. This would appear to indicate that it is the educational or cultural level of the father, rather than his economic level *per se*, which is associated with the child's verbalization of prejudice. In order to maintain simplicity in Table V, the statistically insignificant dichotomies based on father's occupations have been omitted. Ten children in this sample who did not know their father's occupations, or whose fathers were deceased, were excluded from this portion of the analysis.

TABLE VI.—DIFFERENCES IN PREJUDICE VERBALIZATION BY AGE AND GRADE FOR THE NORTHERN SAMPLE (N = 140)

Type of Verbalization	By Age in Years				By School Grade			
	N	Mean Age	σ Age	t Ratio	N	Mean Grade	σ Grade	t Ratio
With Anti-Jewish Prejudice Verbalization	57	8.68	2.15	1.69	57	3.42	2.05	1.74
Without Anti-Jewish Prejudice Verbalization	83	9.39	2.77		83	4.14	2.82	
With Anti-Negro Prejudice Verbalization	49	7.88	2.18	4.49*	49	2.57	2.10	4.87*
Without Anti-Negro Prejudice Verbalization	91	9.81	2.77		91	4.54	2.55	
With Anti-Jewish, Anti-Negro, or Both, Prejudice Verbalization	69	8.17	2.24	3.86*	69	3.16	2.14	3.25*
Without Anti-Jewish, Anti-Negro, or Both, Prejudice Verbalization	71	9.79	2.67		71	4.52	2.75	

* Significant at the one per cent level of confidence.

In dealing with national origin, it was felt that the limited samples available in terms of this variable did not warrant separate statistical treatment for each type of national origin. Instead, and in order to determine whether or not differences in cultural climate associated with national origin might profitably be more closely investigated in the future, children of English, German, Irish and Polish origin were compared as a group with those of all other types of national origin present in the northern sample. This breakdown was used because the majority of children of English, German, Irish and Polish origin showed PV within each of their respective national origin subgroups.

Because of the extremely high incidence of PV in the southern sample, statistical differentiation in terms of variables like age and father's occupation was found unfeasible and was not attempted. While both the northern and southern children were found to express prejudice at the pre-school level, the southern children verbalized prejudice equally at all age levels, while the northern children tended to exhibit less PV with increasing age. Thus, it appears that the results of this study are in keeping with the contention by Ackerman and Jahoda that cultural climate is a determinant of anti-minority attitudes. Their statement that cultural climate in the United States is not consistent seems to be borne out when the northern and southern samples are compared. In the northern sample, the cultural climate of New York City would seem to permit the emergence of such aspects as national origin, age and father's occupation as correlates of PV. On the other hand, the southern sample exhibited PV so uniformly that it may be inferred that a broad southern sub-cultural tendency to PV was operating which obscured the effect of cultural and personal factors found to be associated with PV in the northern sample.

It is interesting to note that the age and school grade differences associated with anti-Jewish verbalization were not significant, while they were for anti-Negro verbalization. A possible reason for this may rest in the difference in level of abstraction required to identify the two minority groups. It would seem plausible that younger white non-Jewish children can, on the basis of skin color, differentiate Negroes from themselves rather easily. On the other hand, ability to recognize Jews as different from themselves rests on the formation of somewhat more abstract concepts such as religious affiliation.

The findings of this study are in contradiction to those of Radke and Sutherland concerning their observation that statements of dislike for Negroes and Jews increased with age. A strict comparison of their results with those of the present study cannot be made because of a number of differences in sampling and techniques used in the two studies.

While it might have been expected that PV differences associated with the sex variable would appear, in view of the difference in level of aggression between male and female to be found in the social rôles assigned to each sex, the findings of this study support a tendency in the direction of higher PV for males which only approaches statistical significance. It is likely, however, that a larger sample would have yielded statistically significant differences.

It should be pointed out that, in assessing the relation of any given variable with PV in this study, no statistical technique was used for holding the effect of other variables constant. Thus, for example, the apparent significant relationship of national origin with PV may disappear when other variables associated with socio-economic status are held constant. In view of this, the results obtained should be cautiously interpreted and regarded primarily as suggesting promising areas for further research.

BIBLIOGRAPHY

1) N. Ackerman and M. Jahoda. *Anti-Semitism and Emotional Disorder*. New York: Harper and Brothers, 1950.

2) T. W. Adorno, et al. *The Authoritarian Personality*. New York: Harper and Brothers, 1950.

3) E. Frenkel-Brunswik. "A study of prejudice in children." *Hum. Relat.*, 1948, 1, 295–306.

4) E. Hartley. *Problems in Prejudice*. New York: Kings Crown Press, 1946.

5) J. B. Holmes. "A study of racial attitudes found in four colleges, including a scale to measure these attitudes." *J. Negro Educ.*, 1943, 12, 69–75.

6) H. Meltzer. "Group differences in nationality and race preferences of children." *Sociometry*, II, 1939, 86–105.

7) M. V. Radke, and J. Sutherland. "Children's concepts and attitudes about minority and majority American groups." *J. Educ. Psych.*, 1949, 40, 449–468.

8) G. Milton Smith. *A Simplified Guide to Statistics*. New York: Rinehart & Company, 1946.

RELATIVE EFFECTIVENESS OF TWO TYPES OF RESPONSE TO ITEMS OF A SCALE ON ATTITUDES TOWARD EDUCATION

CHARLES O. NEIDT

University of Nebraska*

and

WILLIAM R. MERRILL

Iowa State College

One of the problems in constructing a scale for attitude measurement is the selection of the type of response for the items of the scale. Any one of several types of response may be used. For example, the subject may be asked to respond by indicating acceptance or rejection of a single statement, by indicating a position along a continuum of agreement or disagreement for a single statement, or by selecting from two or more statements presented simultaneously that one which most closely corresponds to his point of view.

Before preparing an attitude scale in its final form, some evidence for evaluating various types of response to items is desirable. This evaluation can probably best be accomplished by comparing the results from several administrations of the scale, each involving a different type of response, with some external criterion representing the attitude which the scale purports to measure. From a practical standpoint, however, the number of types of response under consideration will usually be confined to two or three.

It was the purpose of the present investigation to determine the relative effectiveness of two types of item response to a scale on attitudes toward education. The over- or under-achievement of the college students used as subjects was employed as the external criterion for the evaluation. Over- or under-achievement was chosen as the external criterion since it was a form of behavior entirely independent of the scale score, it could be objectively and accurately measured, and it could be assumed to represent an expression of the type of behavior the scale was designed to measure.

---

* The research upon which the results of this article are based was conducted while the senior author was a member of the Psychology Department of the Iowa State College.

METHOD OF PROCEDURE

In an earlier phase of the research project herein described, Dodds[1] constructed two hundred seventy statements of attitudes pertaining to some aspect of education or attendance at college. From an item analysis following administration of these two hundred seventy items to three hundred eighty students, the ninety items which correlated highest with total score were selected for use in the present study.

The Forms of the Scale.—The ninety items finally selected were assembled into two forms of the scale. In one form a five-point response scale of the Likert[3] technique was used as a type of item response. In the other form of the scale, statements of similar content, but expressing different points of view, were paired and the subject was asked to select the statement most indicative of his attitude. For example:

I'd leave school if I____ had a good job offer____	____I intend to finish my college degree ____even if I should get a good job offer.

Since the same statements were used in each form of the scale, the single statement form was composed of ninety items and the paired statement form was composed of forty-five items.

Administration and Subjects.—The two forms of the scale were administered to two hundred one second-quarter freshman students attending the Iowa State College. One-half the subjects took the single-statement form first and one-half took the paired-statement form first. About forty-five minutes was required for administering the two forms.

In addition to responding to the two forms of the scale, the subjects were asked to indicate the number of hours they had studied during the last two weeks. Administration of the scales was accomplished at the end of the fourth week of the quarter.

Scoring.—To construct a scoring key for each form of the scale the responses to the items of the twenty-five highest over-achievers and the twenty-five lowest under-achievers were tabulated. For the construction of the keys over- or under-achievement was defined as the difference between the mean course mark obtained by the student during the quarter in which he took the test and the mean course mark which was predicted for him using as independent variables his American Council on Education Psychological Examination test score and his high-school grade-point average. Over-achievement was indicated by a positive difference (actually larger than expected) and under-achievement by a negative value. In identifying the extreme over- and under-achievers for the item analysis the regression equation was obtained by least squares for the two hundred one subjects and an individual prediction was made for each subject.

The responses to the items in the single-statement form were weighted 4, 3, 2, 1, 0, or 0, 1, 2, 3, 4, according to the direction in which the majority of the over-achievers responded. The paired-statement form was scored 1, 0, or 0, 1, in the same manner. The total score on each scale form was obtained for each subject.

In determining the relative effectiveness of the two types of response, the external criterion was defined as the difference between a subject's actual mean course mark received for the quarter in which he took the scale and the mean course mark predicted for him using his American Council on Education Psychological Examination score and the number of hours studied per week. Although used in constructing the scoring key, high-school grade-point average was dropped from the final analysis since it was felt that the same attitudes toward education were reflected in the collegiate achievement of the subjects. Since the analysis of regression was used in analyzing the results, however, it was not necessary to make individual prediction for each subject.

RESULTS

In Table 1 are shown the correlation coefficients from the final analysis of the five-point single-statement form of the attitudes scale. In Table 2 are shown the correlation coefficients from the final analysis of the paired-statement type of item response.

The split-half reliability coefficients for the two forms of the scale were computed and found to be .67 for the single-statement form and .45 for the paired-statement form. It was also noted by timing a group of subjects that 1.7 times as much administration time was required for the single-statement form as for the paired-statement form.

The correlation between the two forms of the scale was found to be .668.

DISCUSSION

As can be seen from a comparison of the validity coefficients for the two forms of the scale, .355 for the single-statement and .357 for the paired-statement, differences in validity of the two types of response are negligible. From the standpoint of feasibility of administration the paired-statement form would seem to be the more desirable under the circumstances here reported.

TABLE 1.—CORRELATION COEFFICIENTS FOR THE SINGLE-STATEMENT FORM

	Attitude Score (X_0)	ACE Score (X_2)	Hrs. Studied $r(X_3)$*	Hrs. Studied $R(X_4)$†
Course Marks (Y)	.415	.477	.060	.112
Attitude Score (X_0)		.047	.115	.124
ACE Score (X_2)			−.020	.107

$R_{Y(X_2X_3)} = .496$† $R_{Y(X_0X_2X_3)} = .584$† $r_{YX_0 \cdot X_2X_3} = .355$†

* Computed assuming a linear relationship for X_3

† Computed assuming a curvilinear relationship for X_3

TABLE 2.—CORRELATION COEFFICIENTS FOR THE PAIRED-STATEMENT FORM

	Attitude Score (X_1)	ACE Score (X_2)	Hrs. Studied $r(X_3)$*	Hrs. Studied $R(X_4)$†
Course Marks (Y)	.459	.477	.060	.112
Attitude Score (X_1)		.063	−.005	.024
ACE Score (X_2)			−.020	.107

$R_{Y(X_2X_3)} = .496$ $R_{(YX_1X_2X_3)} = .584$ $r_{YX_1 \cdot X_2X_3} = .357$

* Computed assuming a linear relationship for X_3

† Computed assuming a curvilinear relationship for X_3

The relatively high correlation between the two forms (.668) in comparison with their reliabilities (.67 and .45) suggests that the items in this scale are heterogeneous with respect to the criterion. It is conceivable that subscores for the scale could be used more effectively than a single score. It may be that the heterogeneity of the items tended to favor the paired-statement form of response.

As was indicated in the discussion of the scoring key, the item responses of one-fourth the subjects involved in the final analysis constituted the basis for constructing the key. Although only fifty of the two hundred one subjects were involved, and although high-school grade-point average was not an independent variable in the final analysis, the key construction procedure undoubtedly resulted in some spuriousness of the final partial correlations. There is little reason to assume, however, that the key construction procedure affected one type of response more than the other. A further study is being planned to use the key here described in scoring the scales of another group of subjects for the purpose of checking the validity of the scale itself.

SUMMARY

An attitudes toward education scale was administered in two forms to two hundred one college students. Single statements with a five-point response scale were used in one form and the same statements in paired arrangement were used in the other. Partial correlations between the students' mean course marks and each form of the scale, with ACE scores and number of hours studied per week held constant were found to be .355 for the single-statement form and .357 for the paired-statement form. Reliability coefficients were found to be .67 and .45 for the single- and paired-statement forms, respectively.

Although the paired-statement form of response yielded a low reliability coefficient in this investigation, its effectiveness in terms of correlation with an external criterion suggests that it be subjected to further research and be given more consideration in the construction of attitude scales.

SELECTED REFERENCES

1) Barbara Degen Dodds, *Relative effectiveness of paired statements and single statements in a scale designed to measure attitudes toward education.* Unpublished M. S. Thesis. Ames, Iowa, Iowa State College Library, 1949.

2) Daniel H. Kulp, "The form of statements in attitude tests," *Sociology and Social Research.* 18: 18–25. 1933.

3) Rensis Likert, "A technique for the measurement of attitude." *Archives of Psychology.* No. 140, 1932.

4) William A. Owens, "Items form and 'false-positive' response on a neurotic inventory." *Journal of Clinical Psychology.* 3: 264–269. 1947.

BOOK REVIEWS

University of Chicago, Faculty of the College. *The Idea and Practice of General Education. An Account of the College of the University of Chicago.* Chicago: University of Chicago Press, 1950, pp. 333. $3.50.

Colleges in the Nineteenth Century showed a high degree of similarity in their course offerings. Subjects were limited in number, and regardless of the student's postgraduation goals he was expected to follow the same course of study. The introduction of the elective system was the beginning of an accelerating proliferation of courses and departments. While the elective system facilitated the training of specialists in many fields it limited the possibilities of a broad basic education which was claimed to be an advantage of the earlier classical liberal arts. By the end of the first quarter of the Twentieth Century there was concern over the trend toward specialization. Return to the classical program was impractical, so other ways of providing a broad education were sought. Literature of the past twenty-five years reports many experiments devoted to the development of 'general education' programs. The function served by such programs is the same as that of the classical liberal arts, but the content has been changed to conform more closely to the demands of our modern society.

The College at the University of Chicago was one of the first institutions to design a general education program which was consciously experimental. Introduced first in 1931, this program has had many changes in content and method; even today it is not considered to be in its final form. In the volume being reviewed seventeen faculty members who have actively participated in the program describe its growth and its detailed pattern at the present time. The first chapter discusses the problem of general education as it has been considered in the College. The second chapter presents a history of the University of Chicago with special emphasis on the position of the undergraduate College. Part II of the book includes chapters devoted to the curriculum in the three main divisions, viz., humanities, social sciences, and natural sciences, and in the auxiliary areas of mathematics, writing, languages, and integration. In Part III, three chapters on methods discuss teaching, examining, and student advising.

Admission to the College, on the basis of examination rather than completed high-school courses, normally takes place after two high-school years. In each of the first three years the student takes a special integrated course in the humanities, in social science, and in natural science. In addition English, mathematics, and a foreign language are included in the first to third years, respectively. The fourth year is devoted to history and a course designed to integrate the work of the first three years which is entitled "Observation, Interpretation, and Integration." Content as such is not considered the primary function of the three area courses; rather the objective is to develop judgment and methods of analysis and thinking. The B.A. degree is granted on the basis of successful completion of comprehensive examinations rather than on the basis of an accumulation of course credits.

The enthusiasm and clarity with which the authors in the present volume report the work in their several areas should weaken the arguments of even extreme antagonists of general education. The twenty-year experiment at the University of Chicago is as important to the history of higher education in this country as was President Eliot's introduction of the elective system at Harvard over half a century ago. Whether the complete program of the University of Chicago could be introduced in other universities might be questioned; that it affords a plan by which other universities might be guided in adapting general education to their special circumstances cannot be.

C. M. LOUTTIT

University of Illinois

GERALD R. PASCAL AND BARBARA J. SUTTELL. *The Bender-Gestalt Test: Quantification and Validity for Adults.* New York: Grune and Stratton, 1951.

The Bender-Gestalt is one of the performance tasks often used by clinicians to evaluate mental functioning. The subject is required to copy a set of eight simple designs. Until the publication of this manual, the test has been inadequately supplied with scoring standards and norms, and validation has been much needed. Pascal and Suttell administered the tests to hundreds of cases and identified features most likely to occur in records of patients as contrasted with normals. These features are embodied in a careful set of scoring guides of a near-objective sort. Unfortunately the report on item validity (Table 1) involves faulty statistical methods, but the total scoring method appears acceptable. Two scorers use it independently, on a sample of very wide range, with a correlation between scorings of .90.

For norms and validity studies, the investigators rely on hit-and-miss samples, testing whatever groups could be obtained. Granting that availability is a requisite, it is nonetheless unfortunate that so much effort should be expended on unplanned samples in this and other clinical research. The authors present their findings with due caution and with commendable modesty; they often admit to having no idea what certain findings mean where less restrained clinicians would write several pages of rationalization. Yet they are not constricted: at one point (p. 23) they squeeze a hypothesis out of data from three cases.

The evidence for the Bender-Gestalt test is seemingly solid. Not only does it distinguish patients from community persons with bi-serial validities about .70; much more valuable in practice, it discriminates substantially between patients who improve and those who do not, tested at admission. The authors discuss the qualitative indications that their experience suggests consideration of, illustrate records of children, organics, and psychogenic cases with diagnostic discussion, and provide samples for training scorers. There are one hundred four pages of text, apart from appendices, which will be of great assistance to those who use the test. Both the style and organization of the material are rambling, but the authors' ideas advance the usefulness of the Bender-Gestalt appreciably.

LEE J. CRONBACH

University of Illinois

BRUNO BETTELHEIM. *Love Is not Enough.* Glencoe, Illinois: The Free Press, 1950, pp. 386.

Love cannot be given by parents who do not possess it. Parents who do love their children do not have to be told to do so. It is not enough to do the right thing at the right moment. Without the feelings and emotions that go with the act the effect is not educationally sound. For example, in bed-wetting, if the permissive mother also kicks up a fuss about the labor of washing her sheets the child may suffer even greater guilt feelings than the one who is scolded for wetting himself. The foregoing statements illustrate the attitude which underlies both the selection and interpretation of materials in Bettelheim's volume, *Love is not Enough*. The educational philosophy is that of John Dewey and the practice in line with the work of August Aichorn and Fritz Redl. A well told story of how Bettelheim and his staff have tried to apply these principles and practices in their work at the University of Chicago's Sonia Shankman Orthogenic School is what the book is about. It is the story of how an educational institution for emotionally disturbed children sets the stage and permits children who have been mishandled by parents to develop the courage to live and learn through living so that life is again made livable for them. Everyday activities are conceived of as carriers of personal relationships, experiences of mastering tasks previously avoided, and those in which the child in the past experienced defeat.

The chapter captions reveal the nature of everyday experiences illustrated and considered. Illustrative chapter captions are "The Children," "First Encounter," "Events in Sequence," "From Dreams to Waking," "The Challenge of Learning," "Rest and Play," and "The World Outside." The book also contains sixteen illustrations by Myron Davis, the photographer, who lived with the children before doing his work and was in that sense considered by the author a participant observer which is the manner in which he considers the staff.

Although the education at the school is psychoanalytically slanted in its interpretation of behavior, Bettelheim is too realistic an educator to apply orthodox Freudian techniques in dealing with the children, and instead of emphasizing the bringing to light of unconscious he favors techniques for ego bolstering on the spot. Living spot therapy in a sense is what is practiced in this school under his direction. For example, what they try to do in a case of cleanliness is deal with anxieties around cleanliness in the settings in which they arose, that is, in the bathroom and the toilet. By the same token, oral disturbances can be dealt with much more directly in the dining room than in the treatment room; and so on round the clock. The style of interview that is most frequently done by the staff members, called participant observers, and the children is labelled by the author 'marginal interviews.' This is interpretive in character but does not need to interfere with the momentary activity of the group or individual. The purpose may be to clear up an anxiety that interferes with the enjoyment or participation in an activity, or it may be to warn the child of an unavoidable outcome of his behavior that he does not seem to foresee. Again the author does not depend on symbolic thinking or its interpretation for the very good reason that "we live with the children twenty-four hours a day. We do not have to rely on symbolic substitutes for events that took place in the past to be able to understand what may have happened at that time."

Intended chiefly as a description of the way the author's school deals with emotionally disturbed children the volume will specifically serve other specialists in the same filed, but it can also serve to advantage to enlighten individuals concerned with the specifics of child growth and education for growth.

H. MELTZER

Psychological Service Center
St. Louis, Missouri

HADLEY CANTRIL, Editor, AND MILDRED STRUNK. *Public Opinion, 1935–46.* Princeton: Princeton University Press, 1951, pp. lix + 1191.

This book, the first volume of its kind, "aims to present as many public opinion polls as possible in a convenient and useful form." It includes reports from twenty-three organizations in sixteen different countries, and, according to the jacket, it will be indispensable to "historians, sociologists, political scientists, economists, editors, government officials, policy makers, and a host of others concerned with public reaction to current events."

The problem of number of subjects, 'size of samples,' is dealt with in the Introduction in general statements rather than by stating the sizes of samples for each report in the body of the book. The variability is large. The U. S. polls seem to use most commonly, 3,000 to a somewhat larger number; Fortune's

Executive Forecast, 10,000; the National Opinion Research Center, 2,500 to 3,500, but in some cases as many as 7,000; state election polls, 15,000 to 20,000 (in 1936, 30,000). Numbers for foreign polls seem to run from 2,000 to 3,600 with a few exceptions where the samples are as small as 1,000. The telegraphic surveys use generally fewer than 1,000. The important matter is not how many ballots are sent out or how many interviews are planned, but on how many returns are data actually made up. The present reviewer has made special inquiry on this point which does not seem to be clear in the book. It appears that the numbers of subjects, size of samples, would not be less by more than 100 to 150, than the number indicated in the Introduction (pp. viii–ix). The minimum for any inclusion would be determined in terms of the particular country involved, its sampling methods and would be checked by expert authorities.

The contents of the volume are not limited to what many people understand as opinion, but to a much wider concept, including answers to such questions as, "Did you have any burst pipes at home during the recent cold spell?," "Do you ever wear overshoes, rubbers, or rubber boots?," "Have you read . . . ?," "Can you recall having seen . . . ,?" "What is your guess . . .?" and questions eliciting likes and dislikes, preferences, etc. Many, of course, fall under what is generally understood as being matters of opinion: "Which of these plans do you favor . . .?," "Should America have the biggest navy in the world?," "Do you think school teachers should be allowed to spank disobedient children at school?." Although vast in its comprehensiveness, the volume can hardly be thought to fulfill the advertising statement that it includes "every subject under the sun."

The Table of Contents is an index (pp. xv + lix). In such an immense work an index is more than usually difficult. Apparently they are always difficult since there are so few good ones. Despite all the work that has been done on this index, the reader will wish that more page references had been given instead of so many statements, See so and so. One may be obliged to read forty out of eighty reports before he finds the study he desires; a page reference would have saved all of this work. In many cases it is obvious that the omission of page references is justifiable, but omission on so lavish a scale is at least to be deplored.

The immense sweep of subjects from A. A. A. (See Agricultural Adjustment Act) to Yugoslavia makes it impossible to comment in detail upon the contents of the volume. In certain studies it is possible to trace changes in opinion through a number of years. In many reports one can find differences in opinion in terms of age, sex, economic condition, and of known causes such as war, etc. One can look for a special subject which does not appear. Psychiatry does not appear in the index, but there is a strange and interesting heading, Mental Physiology, (at least not new, Carpenter's book entitled *Mental Physiology* was published in 1874). Studies vary in number under a given title from one to nearly one hundred.

Among the topics having the largest number of reported studies are included agriculture, air, military problems, Canada, children and schools, economic problems, education, elections, finance, food, Great Britain, industry, international coöperation, labor, liquor, national affairs, taxation, wages, war and United States. For many studies there are one or a few reports.

There are no catchwords at the tops of pages but the indexing of the book as a whole makes it possible without undue loss of time to run down most but not all subjects.

It is not the province of the reviewer (at least of this one), to raise the question as to accuracy and reliability of data reported in the volume. This has been done at great length and literature on the subject is easily available. It is very pertinent to raise the question as to the influence of this book. It will probably be a sort of bible for large numbers of people for many years. Now, especially, when daily papers and periodicals are constantly quoting results of polls, this volume might become one of the most powerful influences available. For those who wish to vote with the majority, and for those who wish to agree with those who vote with the majority, here is a source which one may use, or because of which one may be unduly influenced.

In this connection it is to be noted that although the Copyright 1951 is printed conspicuously, just as plainly beneath the copyright notice is the following: "The facts recorded in this volume may be freely used in other publications."

The volume is well and strongly bound and should withstand much continual reference to which it will undoubtedly be subjected. The type is clear and easily read. The presentation of studies is such that results can be quickly understood. The printing is in two columns. It is the authority in its field. It will be considered indispensable in a number of different fields.

A. S. EDWARDS

University of Georgia

NORMAN L. MUNN. *Handbook of Psychological Research on the Rat.* Boston: Houghton Mifflin Co., 1950, pp. 598. $7.50.

This work is put forward not as a text but as a complete review of the psychological research on the rat. Even so, it is difficult to see how it can escape being used as a basic text in its field. The comprehensiveness of the book is indicated by the fact that a little more than a hundred pages are required for the bibliography.

The problems, methodology, findings, and interpretations advanced in the book are those the research workers themselves have been concerned with. In all Munn seems to have been interested no more nor no less with systematic and theoretical problems than the men whose work is presented. A nice historical balance has been maintained. Morgan, Small, Washburn, Thorndike, Watson and Carr and other pioneer workers have their day as well as Lashley, Hull, and Tolman, their many students, and others of the present-day.

Undoubtedly psychological research on the rat has contributed as much to the development of psychology in general as that conducted in any other field. This book does not emphasize this contribution. It treats primarily the psychology of the rat. It is significant for what it tells us about the rat. It should be of inestimable value to anyone interested, from any standpoint, in psychological research on the rat. It is doubtful if a work in experimental psychology superior to this, at least in terms of instructional value to the student, is to be found in any field of psychology. Of course it must be admitted that the author has chosen a field that has lent itself to rigorous experimental investigation more readily than most others in psychology.

Substantial progress has been made in all aspects of the field since the appearance of Munn's *An Introduction to Animal Psychology* in 1933, which was actually a work on the psychology of the rat. Perhaps the most striking new developments are to be seen in abnormal and social behavior of the rat and research in systematic psychology.

The ten chapters in the book bear the following titles: 1)

Introduction; 2) Unlearned Behavior; 3) General Activity; 4) Motives, Emotions, and Hoarding; 5) Sensory Processes; 6) The Rôle of Sensory Processes in Maze Behavior; 7) The Learning Process; 8) Some Aspects and Conditions of Learning; 9) Systematic Psychology; 10) Abnormal and Social Behavior.

In reading this comprehensive exposition of psychological research on the rat one is struck with the similarity between the psychology of the rat and the psychology of man. Naturally, the rat is of somewhat limited power. But as far as he goes he is strikingly similar to man. One would be rather hard put to find important psychological phenomena in the rat that are divergent from such phenomena found in man. Psychologists have been cautious about generalizing from animal data. Perhaps our first concern should be to make sure we understand the animal. This is certainly the first concern of the animal psychologist.

J. B. Stroud

State University of Iowa

Werner Wolff. *Diagrams of the Unconscious, Handwriting and Personality in Measurement, Experiment and Analysis.* New York: Grune and Stratton, 1948, pp. 423.

Handwriting as an index of personality has typically received unfavorable and critical treatment in psychological literature. The reason for this, according to the author of the volume under review, is that the experiments have been limited in scope and have merely proved the absence of arbitrary relationships between single graphic signs and single personality traits. This, he contends, does not disprove the possible relationship between the total graphic expression and total personality. Accordingly, in this comprehensive volume of handwriting and personality, measurement, experiment, and analysis, the author, well known for his experimental work in depth psychology, attempts to evaluate and illustrate how handwriting can be best subjected to experimental procedure in the light of a more dynamic psychology. The method he advocates and applies is to select characteristic themes of personality and relate these to analysis of handwriting. A thematic approach is in line with modern trends in projective techniques. It is the author's impression that direct analysis of expressive movements, particularly handwriting, without intermediary materials for projection allows more precise and more defined application of scientific criteria than projective methods. The volume as a whole is impressive appearing and has the surface earmarks of a thoroughgoing and scholarly comprehensive work. Evaluative considerations, analysis of hypotheses, plentiful illustrations, lists of specimens of handwriting, attempted correlation with biographical material, graphs, drawings, tables, an extensive bibliography and index are all included. Does a more thoroughgoing analysis of the material selected, its organization and emphasis show up as well as its surface manifestations?

The content of the book is presented in two large parts. Part I, entitled "Form and Movement," includes a general treatment of the whole subject, containing the author's thesis and argument for a consideration of handwriting in personality measurement. In the seven chapters of Part I topics considered include symmetry, rhythm, configuration, consistency in personality, diagrams of personality, and the origin of configuration. Part II, called "Expression and Movement," includes chapters on unity of expressiveness, empathy and projection, symbolization, interpretation of graphic expression, the author's evaluation of the self, analysis of indicators of graphic expression, and a final chapter on handwriting and biography. The author contends that the findings he reports in this volume indicate that the unconscious movements in writing represent a reign of order, proportion, and configuration appearing in the same way as if they had been consciously calculated, measured, and constructed. Statistical evidence for claims made are presented. The matching technique of which the author is one of the founders is applied here by him to show a more meaningful relationship and connection between handwriting and other forms of personality expression. This is in line with his general feeling that graphic expression is not only patterned by some factors of personality but by the unity of personality.

In spite of the author's attempts to show consistency of treatment of the various phases of the topic he considers, occasional inconsistencies and over-emphasis appear. In Part I, particularly, and to some measure through the book, he emphasizes a need for dynamic and configurational approaches. However, in Table 16 he lists the main characteristics of graphic expression indicative of a relationship between the appearance of writing and characteristics they signify without much reservation. In his chapter on biography he analyzes the hardwriting of Lincoln, Spinoza, Kant, and others with biographical materials. In Lincoln's graphic expression, we are told inhibitory elements seem to dominate. "Passivity, restriction, negativism, psychobiological disturbances, dreaminess, introversion and sensitivity" are said to be manifested. But, we are also told that the opposite elements of activity and determination are also present in Lincoln. The graphic expression indicates this potentiality although it does not show that Lincoln made his own passivity an element of activity and turned his melancholy into power. This final statement is an interpretation in line with biographical data, but it is not revealed in the analysis of handwriting. An analysis of Lincoln's personality on the basis of handwriting only would certainly be an incomplete one and, in terms of neglecting the significant, a false one. Similar criticism can be made of other attempts at correlating biographical data with handwriting. That this material represents real clinical validation is more questionable than the author seems to imply or express.

The author is a serious student, interested in experimental approaches to depth psychology and he is serious in this effort of his. What he claims for handwriting is worthy of consideration for further investigation, however, rather than for uncritical acceptance. The fact that some of the conclusions the author arrives at are best classified as based on subjective judgment rather than real clinical validation does not disprove the value of the contribution. The least one can say is that the volume does have in it materials that are subject to interesting discussion and suggestive of experimental procedure for further investigation of some of the problems raised, even though they do not represent a series of conclusions concerning handwriting in personality measurement that represent validated insights for personality understanding.

H. Meltzer

Psychological Service Center
St. Louis, Missouri

C. P. Stone, Editor. *Annual Review of Psychology*. Vol. I. Stanford, California, Annual Review, Inc., 1950, pp. 330.

In 1950 Annual Reviews, Inc., has added four fields to its listing: medicine, plant physiology, physical chemistry, and psychology. The oldest, *Annual Review of Bio-Chemistry*, is now in its eighteenth volume. *Annual Review of Psychology* is in its eleventh volume.

The chapters in *Annual Review of Psychology*, eighteen in all, cover most of the major topics found in the Contents of the *Psychological Abstracts*. The *Review* will not compete with the *Abstracts*, but will supplement it in various significant ways. In the *Review* the authors have organized, synthesized and critically evaluated the research and theoretical writing of the year covered in their respective fields. (The period covered is roughly July 1948–July 1949.) In many cases the authors have related the year's development to recent previous work. This publication should aid psychologists greatly in maintaining contact with fields of psychology that lie outside their own specialization. It should be very popular with graduate students.

The authors in this first volume have set a high standard. There are several outstanding chapters. Among these, those of Melton, Sears, Cronbach, and Thorndike should be of special interest to educational psychologists.

Authors and Chapter titles are as follows: H. E. Jones and N. Bayley, Growth, Development and Decline; A. W. Melton, Learning; N. R. Bartlett, Vision; E. B. Newman, Hearing; F. A. Geldard, Somesthesis and the Chemical Senses; R. L. Thorndike, Individual Differences; R. R. Sears, Personality; J. S. Bruner, Social Psychology and Group Processes; C. L. Shartle, Industrial Psychology; D. O. Hebb, Animal and Physiological; N. A. Cameron, Abnormalities; H. F. Hunt, Clinical Methods: Psychodiagnostics; W. U. Snyder, Clinical Methods: Psychotherapy; L. J. Cronbach, Educational Psychology; R. F. Berdie, Counseling Methods: Diagnostics; E. S. Bordin, Counseling Methods: Therapy; D. A. Grant, Statistical Theory and Research Design; D. M. Johnson, Problem-Solving and Symbolic Processes.

J. B. Stroud

State University of Iowa

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 42 December, 1951 Number 8

PATTERNS IN THE MASS COMMUNICATIONS TASTES OF THE YOUNG AUDIENCE*

PAUL I. LYNESS†

Development Division, The Curtis Publishing Company, Philadelphia

This article summarizes a recent study of the mass communications habits and tastes of young people in school. The major goal of the study was to discover what 'patterns of interest' in the content of the mass media as a whole characterized five different age groups of boys and girls. By 'patterns of interest' we mean here the tendency for certain subject matter to cluster in the content preferences of any given age-sex group, and the tendency for these preferences to carry over from medium to medium.[1]

---

* The study of which this article is a summary was the author's doctoral dissertation in mass communications, written under the direction of Professor Leslie G. Moeller, Director of the School of Journalism, State University of Iowa.

† Formerly head of the Research Division, School of Journalism, State University of Iowa.

[1] Alice Sterner (*Radio, Motion Picture and Reading Interests; A Study of High-School Pupils*, Contributions to Education, No. 932, Teachers College, Columbia University, 1947) has provided us with one approach to this problem of communication interest patterns. She strove to answer these questions: "What really impels youth to spend so much time (in mass communications activities?) . . . Is it the medium itself or a major adolescent interest in adventure, humor or love, which attracts the pupil?" Some other researchers who have studied interests in all or nearly all the media simultaneously are: Wilbur Schramm, "Reading and Listening Patterns of American University Students," *Journ. Quarterly*, March, 1945; Walter J. Scott, *Reading, Film and Radio Tastes of High-School Boys and Girls*, Wellington, New Zealand, 1947; Sidney L. Pressey and Francis P. Robinson, *Psychology and the New Education*, New York, 1944, pp. 126–139; Frederick J. Meine, "Radio and Press among Young People," in Lazarsfeld and Stanton, *Radio Research–1941*, New York, 1941, pp. 189–223; Paul A. Witty and Anne Coomer, "Activities and Preferences of a Secondary-

It was felt that if rather detailed patterns of mass communications interest could be established and described meaningfully, editors, station managers, and producers would find them useful in 'beaming' their output to specific audiences of young people. Such information would also enable teachers better to understand their pupils and to utilize the mass media more effectively in their teaching programs. Social researchers, too, preoccupied with the prediction and control of behavior, might find the material valuable in testing the interaction between youthful mass communications interests and personality, adjustment and behavior.

With this as the problem, then, a rather comprehensive study was designed and carried out in the public schools of Des Moines, Iowa, in March, 1950.[2]

METHOD

Samples in five grades—third, fifth, seventh, ninth and eleventh—were selected by essentially random methods, although an initial stratification by sex and school attended was made. Students above the third grade filled out questionnaires in the classroom. Third-grade students were interviewed individually. The questionnaires were administered by teachers in the Des Moines public school system and graduate students at the State University of Iowa, all of whom were carefully briefed on the purposes of the survey and the procedures to be followed. The questions covered the amount and kind of reading, radio listening and movie-going; the preferred types of mass communications content (by medium); the names of preferred newspaper columnists, radio programs, movies, magazines, books and comic books; and preferred media.

A series of six pretests involving pupils in the third, fourth, fifth, seventh, eleventh and twelfth grades permitted a practical evaluation of the study design, the questionnaires and the methods employed.

School Group," *Journal of Educational Psychology*, February, 1943, pp. 65–76.

[2] The study was in large part financed by the *Des Moines Register and Tribune Company*. The writer wishes to thank Henry Kroeger, Director of Research and *The Iowa Poll, Des Moines Register and Tribune*, for his extensive personal assistance in carrying out the project.

CHARACTERISTICS OF THE SAMPLES

The total sample of 1,418 school children was made up of six hundred ninety-one boys and seven hundred twenty-seven girls. Over three hundred students in each of grades V, VII, IX and XI took part in the study. In addition, one hundred fifteen third-grade students, fifty-nine boys and fifty-six girls, were interviewed.

The breakdown of the total, by occupation of father, showed the following distribution: professional, eleven per cent; managers and proprietors, fifteen per cent; clerical and sales, twenty-two per cent; service, six per cent; skilled labor, twenty-two per cent; semi-skilled labor, eighteen per cent; unskilled labor, three per cent; other, three per cent.

By education of father, the percentage breakdown for fifth-grade students was: grade school, 5.6; high school, 26.6; college, 35.6; don't know, 32.2. For seventh-grade students, the breakdown in percentage (following the same order as above) was 4.6, 47.4, 26.1 and 21.9 respectively; and for the ninth- and eleventh-grade respondents, approximately 13.5, 44.3, 28.5, and 13.6. Information on the education of parents was not obtained from third-grade students.

Other background data obtained in the survey included ownership of automobiles, telephones, radios and TV sets; church affiliation and attendance; years of residence in Des Moines; preferred presidential candidate in 1948; possession of a library card; and recreational activities at home and away from home.

REPRESENTATIVES OF THE SAMPLES

The samples were substantially representative of their respective grade populations, in terms of sex and school attended, by virtue of the initial stratification. Figures on parental education and occupation for all public school children were not available. However, an estimate of 'representativeness' in terms of these factors was made as follows: A statistical comparison was made of educational and occupational data gathered by *The Iowa Poll*[3] on city adults with children in school and city adults without children in school (state-wide surveys). The differences between these groups were insignificant at the five per cent level of confidence.

---

[3] *The Iowa Poll* is a weekly, state-wide survey of what Iowans think on leading topics of the day. It is sponsored by *The Des Moines Register and Tribune.*

Then, the educational and occupational data in the present survey were compared with three recent *Iowa Polls* of adults in the city of Des Moines. This test revealed no significant differences in the occupational categories (at the five per cent level of confidence), except in 'Managers and Proprietors,' where the difference was significant at the two per cent level. The present samples contained more people from middle and lower status homes than *The Iowa Poll* samples. On the other hand, the test showed that the present samples were somewhat overloaded with children whose parents had attended high school and college.

THE SURVEY CITY

Des Moines, at the time of the survey, had a population of approximately 176,954 (1950 census). The morning *Register* and the evening *Tribune*, both published by the *Des Moines Register and Tribune Company*, were the city's daily newspapers.

All four major networks had outlets in Des Moines, and, in addition, two independent stations were located there. In 1949, there were twenty-four moving picture theatres in the city. Des Moines citizens had access to over half a million library books in the ten public libraries, the Drake University library and the State Traveling Library.

THE QUESTIONS RELATING TO PREFERRED TYPES OF CONTENT

In the questionnaire given to fifth-, seventh-, ninth- and eleventh-grade pupils, seven questions dealt with preferred types of subject matter in the various communications media. One question was devoted to each of the following: newspapers, radio, moving pictures, types of magazines, magazine stories and features, books, and types of news.

A check list of content categories accompanied each question. The pupils were asked to select and rank their four favorite types of content from each list. Thus, the question relating to motion pictures was phrased as follows: "11. Here are several different types of movies. Which four do you usually like best? Write '1' before the type of movie you like best. Write '2' before your second choice. Write '3' and '4' before your third and fourth choices."

The check list of motion picture categories appeared below the question (See Table III).

The radio and magazine check lists contained several names of programs and magazines exemplifying the categories. The illustrative programs had high Hooperatings in Des Moines at the time of the study and represented the various networks and stations heard in Des Moines. For example, "Western shows like The Lone Ranger, Roy Rogers, Tom Mix, Hopalong Cassidy, etc.," and "Radio columnists like Jimmy Fidler, Louella Parsons, Walter Winchell, Sheila Graham, etc." The magazines named were high circulation magazines in each classification. For example, "Sports magazines like *Sports Afield*, *Outdoor Life*, *Sport*, etc.," and "Girls' magazines like *Senior Prom*, *The American Girl*, etc."

In determining the preferences for each sex-grade group, the first, second, third and fourth choices for each medium were weighted and combined. (First choices were weighted 4; second choices, 3; third choices, 2; and fourth choices, 1.) The weighted scores for each medium were then converted to percentages. This made easier the comparison of preferences in the different media.

FINDINGS

Tables I–VII contain the content ratings for each sex-grade group, and for all media, plus types of news.

It should be borne in mind that these ratings are based upon the weighted first, second, third and fourth choices, and not upon a single favorite in each medium. The four-choice approach was used in order to get some measure of intensity or depth of interest.

A tabulation of first choices alone disclosed high group interest in a few categories. For example, the number one newspaper favorites of about three-fourths of all respondents in all grades were either comics or sports. With only a quarter of the first choices remaining, to be distributed among fifteen to twenty categories, the ranking of the secondary interests became somewhat difficult and imprecise. On the other hand, the tabulation of four weighted choices revealed strong secondary interests in

TABLE I.—FAVORITE PARTS OF THE NEWSPAPER
(In Per cents of Weighted 1st, 2nd, 3rd and 4th Choices)

Boys					Girls			
5th	7th	9th	11th		5th	7th	9th	11th
2.2	2.1	1.4	2.0	Advertisements	.8	1.9	1.0	1.2
.1	.4	.2		Book Reviews	1.0	.2	.5	.1
.6	.7	.3	.3	Business pages	.1	.4		.1
1.4	1.1	2.0	2.9	Columns on public affairs	.9	.9	2.1	2.8
.5	.2	.8	.9	Columns on love & personal problems	3.1	4.8	5.7	4.8
33.5	30.4	24.7	24.8	Comics	34.4	34.7	26.8	22.9
4.8	3.2	1.0	1.1	Continued stories	5.3	4.0	1.1	.3
1.4	.5	2.0	2.2	Editorials	.6	1.2	1.0	2.4
1.2	1.1	1.4	1.1	Gossip Col's	1.0	2.0	3.4	3.0
6.2	6.7	7.8	5.8	Headlines	8.0	5.9	7.1	6.0
.2	.3	.5	.4	Letters to editor	.4	.7	.8	1.5
11.1	10.0	7.8	4.3	Movies or radio news	12.2	13.2	8.8	5.8
4.7	9.5	12.6	15.0	News (general)	7.6	6.0	10.6	10.8
8.6	7.2	7.8	8.0	Pictures	12.1	9.8	7.5	5.7
		.1		Society; Women's features	1.6	3.0	5.6	12.2
22.7	26.4	28.1	30.5	Sports	9.8	10.7	17.0	20.3
.7	.2	1.4	.7	Other	1.0	.7	1.0	.1
.1		.1		None	.1			
1456	1616	1547	1410	Weighted total	1557	1945	1458	1576

such newspaper categories as movie and radio news, women's features, and general news, depending upon grade and sex. Comparisons of first choices with the combined four choices did show considerable correspondence in the direction and magnitude of changes from grade to grade.

One important qualification to the findings contained in Tables I–VII should be considered. The rating of each content category for each sex-grade group is a relative rating; that is, it is relative to all other categories for a given medium. Thus, the fact that interest in newspaper comics fell off somewhat with age, relative to other newspaper categories, does not mean that the older children's absolute interest in comics was less. It merely means that the relative rank of comics among other newspaper categories was lower among the older boys and girls.

TABLE II.—FAVORITE TYPES OF RADIO PROGRAMS
(In Per cents of Weighted 1st, 2nd, 3rd and 4th Choices)

Boys					Girls			
5th	7th	9th	11th		5th	7th	9th	11th
14.2	11.2	4.3	1.7	Adventure serials	9.4	6.5	2.9	1.1
13.7	15.1	19.1	18.1	Comedy shows	13.5	15.4	16.7	13.3
.1	.4	1.1	2.3	Classical music	1.2	1.1	2.1	5.0
.3	.5	.1	.1	Love serials	3.6	1.8	1.8	.8
12.1	18.0	17.3	17.4	Mystery plays	12.0	18.2	16.7	11.7
1.0	2.8	2.6	3.4	News	1.3	.6	2.0	2.6
8.5	4.0	5.3	3.8	Oldtime and west. music	7.7	5.8	2.9	2.6
6.0	7.5	9.6	10.0	Other plays (radio drama)	12.0	18.7	20.3	21.5
.5	2.5	6.8	15.8	Popular music	3.2	7.8	15.1	23.3
8.4	7.6	8.3	4.2	Quiz shows	12.3	10.4	7.9	5.5
.6	.7	1.1	1.5	Radio columnists	.5	.6	1.4	1.8
.4	.6	1.1	1.5	Religious programs	1.2	1.9	2.5	2.1
8.9	12.4	13.7	15.5	Sports	2.7	3.0	4.1	6.5
.2	.1	.3	.2	Talks; Forums		.1		.3
24.8	16.2	9.0	3.6	Western shows	19.2	7.4	3.1	1.1
.3	.4	.3	.9	Other	.2	.7	.5	.8
1573	1632	1574	1379	Weighted total	1581	1930	1470	1587

Several questions in the study were designed to get from the boys and girls information on what they actually read, heard on the radio and saw in the movies. The responses to these questions indicated a sizeable increase in the variety of reading, listening and viewing from the third through the eleventh grades. For instance, analysis of 'regularly read' newspaper content showed an increased or stable proportion of responses, from grade to grade, in every category except 'book reviews' and 'continued stories.' Coming back to the example above, between 95 and 100 per cent of the boys and girls in all grades said they read the comics regularly.

TABLE III.—FAVORITE TYPES OF MOTION PICTURES
(In Per cents of Weighted 1st, 2nd, 3rd and 4th Choices)

Boys					Girls			
5th	7th	9th	11th		5th	7th	9th	11th
11.1	14.8	18.4	19.6	Adventure	9.2	9.6	9.8	9.4
14.4	8.0	5.4	3.7	Cartoons	16.8	6.9	3.0	1.9
14.7	15.5	13.5	15.2	Comedy features	14.5	13.8	12.7	9.7
2.1	4.3	4.1	6.2	Historical movies	.7	.7	3.0	5.0
.7	4.9	5.1	2.7	Horror movies	1.0	3.7	3.2	2.0
1.3	1.0	2.2	2.3	Love and romance	8.0	12.9	12.8	14.8
2.2	1.7	6.9	11.7	Musicals	11.7	18.1	22.2	28.4
9.1	10.3	12.4	11.5	Mystery	9.5	10.3	13.5	8.2
.4	1.0	2.2	1.2	Religious movies	1.9	3.2	4.9	4.3
.3	.3	.8	1.5	Social problems	.8	.7	1.2	3.5
14.8	17.5	15.5	13.7	War	1.8	2.6	3.6	4.5
28.2	18.1	12.2	8.9	Westerns	18.6	11.8	5.9	2.5
.7	1.8	1.2	1.4	World of make believe	4.4	4.5	3.0	3.4
	.8	.1	.4	Other drama	1.1	1.2	1.2	2.4
1339	1567	1559	1340	Weighted total	1365	1877	1457	1519

In radio program types listened to regularly, there was an increased or stable proportion of responses, from grade to grade, in all categories, except quiz shows, westerns, adventure serials, old-time and western music, and love serials.

Keeping in mind these reservations, one may make three general statements with respect to the data contained in Tables I–VII. 1) Both grade and sex influenced the degree of preference for certain types of content. 2) Preferences for certain types of content tended to remain at about the same level from age group to age group. 3) The tables disclose at least three hypothetical 'patterns of interest' or generalized interests, which may be described as (a) violence and adventure; (b) love, private life, and glamor; and (c) 'educational' content (content which may be described as primarily informative and serious).

TABLE IV.—FAVORITE TYPES OF MAGAZINES
(In Per cents of Weighted 1st, 2nd, 3rd and 4th Choices)

Boys					Girls			
5th	7th	9th	11th		5th	7th	9th	11th
3.7	3.8	4.6	1.3	Adventure	2.0	1.6	.6	.1
12.4	15.8	12.8	6.2	Boys'	.4	.1	.3	.1
9.7	8.8	6.1	3.7	Detective story	7.6	5.3	3.0	.4
4.0	2.7	2.4	1.4	Farm	2.2	.3	.3	.5
2.6	3.9	11.7	14.4	General appeal	4.4	7.3	11.9	13.5
.5	.2	.4		Girls'	11.4	16.6	15.7	15.1
1.5	.9	.9	1.8	Home service	6.2	4.6	3.3	6.3
	.4	.1	.1	Literary	.1		.2	
4.1	4.9	5.6	9.0	Men's	.8	.4	.1	.7
3.4	3.6	2.8	1.6	Movie	12.6	16.4	14.3	9.0
1.1	1.4	1.3	2.8	News	.7	.5	.2	.9
5.5	7.8	9.5	16.5	Picture	9.4	7.9	9.2	11.2
8.1	9.4	12.4	14.1	Science	1.9	.6	.5	.2
19.1	20.6	21.7	23.3	Sports	6.8	3.2	3.0	2.3
.8	1.3	1.5	1.2	Religious	1.2	1.8	2.0	1.2
1.0	1.1	1.2	1.0	Romance	8.5	12.3	11.5	7.2
22.0	13.0	4.8	1.6	Western	13.5	5.8	2.3	.3
.5	.1			Women's	8.2	8.4	9.1	10.3
	.3	.2		Young women's	2.1	6.9	12.5	20.7
1312	1562	1560	1364	Weighted total	1354	1887	1441	1580

I. INFLUENCE OF GRADE AND SEX ON PREFERENCES FOR CERTAIN TYPES OF CONTENT

Those effects of grade and sex which were especially marked will be discussed under two major headings:

A) Categories which rose in relative popularity with the grade level of the respondents:

Both sexes.—The preference level of both boys and girls rose steadily, from the fifth through the eleventh grades, in the case of sports and general news in the newspaper; radio drama, popular music and sports on the radio; movie musicals and historical movies; general magazines; sports stories and features in magazines; and international and foreign news in types of news.

Popular music, radio drama, and movie musicals were consistently more popular with the girls than the boys. Newspaper sports and general news, radio and magazine sports, and international news were consistently more popular with boys than girls.

Boys.—The boys' preference level, but not the girls', increased with respect to the following: movie adventure; science and invention in magazines, books, and types of news; men's magazines; picture magazines; adventure books and books on sports. All these categories were more popular with boys than girls, except for picture magazines and adventure books which were about equally popular with the sexes.

Girls.—The girls' preference level, but not the boys', increased with respect to the following: women's features and society, and gossip columns in the newspaper; classical music on the radio; love and romance, religion, and war in the movies; love and romance, and women's subjects in magazines; magazines for women and young women; historical fiction, and love and romance in books. All these categories, except war and historical fiction, were more popular with girls than boys.

B) Categories which fell off in relative popularity as the grade level of the respondents rose:

Both sexes.—The preference level of both boys and girls dropped, from the fifth through the eleventh grades, in the case of comics, movie and radio news in the newspaper; westerns in radio, movies, magazines and books; adventure serials, old-time and western music, and quiz shows on the radio; movie cartoons; detective story magazines; and auto wrecks and Hollywood in types of news. Westerns and adventure serials were consistently more popular with boys than girls. Movie and radio news was consistently more popular with the girls.

Boys.—The boys' preference level, but not the girls', fell off with respect to boys' magazines and movie magazines; and crime

TABLE V.—FAVORITE TYPES OF MAGAZINE STORIES AND FEATURES

(In Per cents of Weighted 1st, 2nd, 3rd and 4th Choices)

Boys					Girls			
5th	7th	9th	11th		5th	7th	9th	11th
15.2	13.3	13.6	17.2	Adventure	12.4	9.4	8.6	6.1
2.2	1.1	.5	1.4	Art and music	6.1	2.5	2.3	3.0
.4	.3	.2	.6	Business and industry	.1	.1		.1
6.1	7.4	7.8	5.0	Cartoons	8.2	8.6	6.7	5.4
8.4	8.1	9.0	6.7	Crime; Police; gangsters	4.9	4.6	4.7	1.2
5.4	3.3	4.5	5.0	History and hist'l fiction	1.2	1.2	2.1	3.9
2.6	2.2	1.2	1.8	Hollywood and movies	8.9	10.7	11.3	7.6
.4	.1	.1	.4	Home service	.4	1.1	.1	1.4
10.1	11.8	12.5	10.6	Humor and comedy	13.9	15.9	16.6	17.3
	.3	.5	.2	Letters to editor	.4	.6	1.6	.7
.8	.9	2.0	.9	Love and romance	6.6	13.5	13.9	18.8
10.1	9.8	7.8	8.0	Mystery; detective	8.2	9.6	9.0	4.4
.1	1.0	1.5	1.4	Public affairs		.1	.2	1.3
.3	1.2	1.6	.9	Religion and churches	2.7	2.0	2.9	2.1
2.8	4.3	8.4	9.2	Science and invention	1.3	1.4	1.7	1.3
.6	.5	1.3	.4	Social problems		.1	.7	2.4
10.5	16.1	17.8	22.9	Sports	3.7	4.1	5.6	8.9
.2	1.7	1.4	2.4	Travel	1.0	.7	1.2	2.7
22.4	13.9	7.0	4.1	Westerns	14.7	6.1	2.3	1.2
	.1			Women's subjects	1.6	5.0	7.0	9.0
1.4	2.0	.4	.2	World of make believe	3.7	2.4	1.5	1.0
	.6	.9	.7	Other		.3		.2
1242	1556	1519	1334	Weighted total	1359	1790	1454	1557

in types of news. Boys' magazines and crime news were more popular with boys than girls. Movie magazines were more popular with girls than boys.

Table VI.—Favorite Types of Books
(In Per cents of Weighted 1st, 2nd, 3rd and 4th Choices)

Boys					Girls			
5th	7th	9th	11th		5th	7th	9th	11th
12.8	15.1	19.2	18.3	Adventure	14.6	16.2	14.9	15.8
2.7	1.1	.6	1.5	Art or music	8.5	3.0	2.1	3.2
.8	.9	2.2	1.2	Biography	2.7	1.4	2.1	3.9
6.1	4.1	5.0	5.5	History and hist'l fiction	1.0	2.0	3.6	9.2
.2	.4	.5	.1	Home service	.6	1.0	1.7	1.4
13.4	11.8	12.0	13.0	Humor and comedy	14.8	20.4	21.0	20.6
.5	.7	1.5	1.2	Love and romance	8.5	13.6	15.6	18.1
9.8	10.2	8.8	7.5	Mystery; detective	10.1	13.7	13.1	5.9
.2	.3	.8	.9	Public affairs	.1	.1	.1	
	.9	1.0	2.2	Religious books	2.1	2.6	3.7	3.2
3.9	5.6	8.3	7.1	Science and invention	3.0	2.7	1.6	1.6
.2	.2	.1		Social problems	.1	.2	.8	1.7
13.1	16.8	19.2	22.0	Sports	7.1	5.6	8.5	6.9
.6	1.8	1.3	1.9	Travel	1.9	1.2	1.9	1.9
11.1	12.2	8.0	11.4	War	1.2	1.2	.9	1.5
22.1	14.7	8.3	5.0	Westerns	17.3	8.1	4.0	2.0
2.1	1.6	1.3	.4	World of make believe	6.1	4.8	2.6	2.3
.4	1.6	1.9	.8	Other	.3	2.2	1.8	.8
1319	1595	1561	1391	Weighted total	1429	1882	1489	1566

Girls.—The girls' preference level, but not the boys', fell off with respect to magazine adventure; the world of make believe in magazines and books; and sports magazines. Magazine adventure and sports magazines were more popular with boys than girls. The world of make believe was more popular with girls than boys.

TABLE VII.—FAVORITE TYPES OF NEWS
(In Per cents of Weighted 1st, 2nd, 3rd and 4th Choices)

Boys					Girls			
5th	7th	9th	11th		5th	7th	9th	11th
16.3	13.6	8.9	6.9	Auto wrecks; accidents	9.3	13.4	9.2	6.4
5.4	1.9	1.2	1.6	Art, lit., and music	12.0	5.4	3.8	5.9
.1	.3	1.1	1.0	Business and industry	.2	.1	.7	.6
18.9	17.3	16.3	13.6	Crime; police; gangsters	10.8	11.1	12.6	8.9
8.1	9.8	8.6	8.4	Des Moines news	13.0	10.5	12.9	14.5
9.6	7.9	8.0	9.1	Disasters	6.9	9.7	10.6	9.2
3.4	7.3	10.4	7.7	Internat'l and foreign news	3.9	3.9	7.0	7.5
.5	.7	2.2	1.5	Labor Prob.	.7	.3	.5	1.1
1.2	.2	.7	.1	Love and marriage problems	5.9	4.3	3.4	4.2
1.6	2.0	3.3	4.2	Military affairs	.2	.6	.3	.3
6.4	5.8	3.0	2.5	Movies and Hollywood	16.8	19.0	14.5	9.6
1.1	.6	2.5	3.3	Politics and gov't	.8	.9	1.2	2.7
1.1	2.2	1.6	1.3	Religion and churches	2.8	3.7	5.1	3.0
4.6	5.9	8.0	10.2	Science and invention	3.1	3.2	1.8	3.7
.8	.7	1.1	.3	Social problems	.9	1.1	1.6	2.9
20.6	23.6	22.8	28.3	Sports	12.7	12.2	14.5	19.1
.3	.2	.3		Other		.6	.3	.4
1368	1609	1577	1388	Weighted total	1438	1909	1480	1580

II. CONTENT PREFERENCES WHICH TENDED TO REMAIN AT ABOUT THE SAME LEVEL FROM AGE GROUP TO AGE GROUP

Those categories showing consistent ratings from grade to grade will be discussed in terms of both sexes, of boys, and of girls.

Both sexes.—The following categories drew high and consistent ratings from boys and girls in all grades: newspaper comics and sports (although the percentage score for comics fell off somewhat while the score for sports rose); comedy in the movies and on the radio; humor in magazines and books; mystery and detective in radio, movies and books. Sports and movie adventure were more popular with boys than girls. Humor in magazines and books was more popular with girls than boys.

A consistent secondary interest was shown in horror movies; crime, cartoons, history and historical fiction and mystery in magazines; disaster, religion, and art-literature-music in types of news.

Consistently low in the preferences of both sexes were advertisements, book reviews, public affairs, and gossip columns, in the newspaper; letters to the editor in newspapers and magazines; business and industry in all media; news, religious programs, talks and columnists on the radio; social problems in all media; serious drama in the movies; farm, literary, and news magazines; travel articles in magazines; labor problems, military affairs, and politics in types of news.

Boys.—The following categories rated consistently high with the boys in all grades: war in movies and books; adventure stories in magazines; and sports magazines. All were more popular with boys than girls. The boys showed a secondary interest in humorous books, adventure and movies magazines, and history in books.

Low on the boys preferred lists were columns on love and personal problems in the newspaper; love and romance in movies, books, and magazines; fantasy, and religion in the movies; magazines for girls, women and young women; home service and romance magazines; art and music in books and magazines; Hollywood in magazines; biography in books; and news of love and marriage problems.

Girls.—The following categories rated consistently high with the girls in all grades: romance, movie, and girls' magazines; humorous books; magazine stories about the movies and Hollywood; and crime news. All were more popular with girls than boys, except crime news.

The girls showed a secondary interest in columns on love and personal problems in the newspaper; religion and fantasy in the movies; home service and picture magazines; books on sports, religion and art or music; biography; art and music in magazines; science and invention in books and types of news; and love and marriage problems in types of news. Historical movies; adventure, science, boys', men's, and science magazines; and war books were low among the girls' preferences at all grade levels.

III. THREE HYPOTHETICAL PATTERNS OF INTEREST

In order to get a more generalized picture of interest tendencies among young people, three hypothetical interest patterns were established, under which specific subject-matter categories could be classified. It was desired to uncover the extent to which these patterns characterized any given grade-sex group.

The three hypothetical patterns were (a) Adventure and Violence; (b) Love, Private Life and Glamor; and (c) 'Educational' content. All categories (in each medium) which were subsumable under these headings were grouped together for analysis.

Mean preference ratings for each pattern were calculated for each grade-sex group by dividing the sum of the ratings for all applicable categories in each and every medium, by the total number of occurrences of the categories in each and every medium. There were twenty-two total occurrences in the adventure and violence classification; thirteen occurrences in the case of love, private life and glamor; and thirty-three occurrences in the case of 'educational' content. These ratings were subject to the characteristic disadvantage of means in that they could be skewed by extreme cases; however, there were few extreme cases in the distributions concerned. Figure I shows graphically the mean ratings of the three hypothetical patterns of interest by grade and sex.

The results indicated marked sex differences in the preferences for adventure and violence, and love, private life and glamor. The boys' ratings exceeded the girls' in every violence and adventure category. The girls' ratings exceeded the boys' in every love, private life and glamor category. Neither sex indicated any substantial preference for 'educational' content.

There were also clearcut differences by grade in the case of adventure and violence, where the relative preference of both boys and girls declined steadily as the grade level rose. Differences by grade in the preference shown love, private life and glamor were not clear cut.

It should be emphasized that most of the individual categories

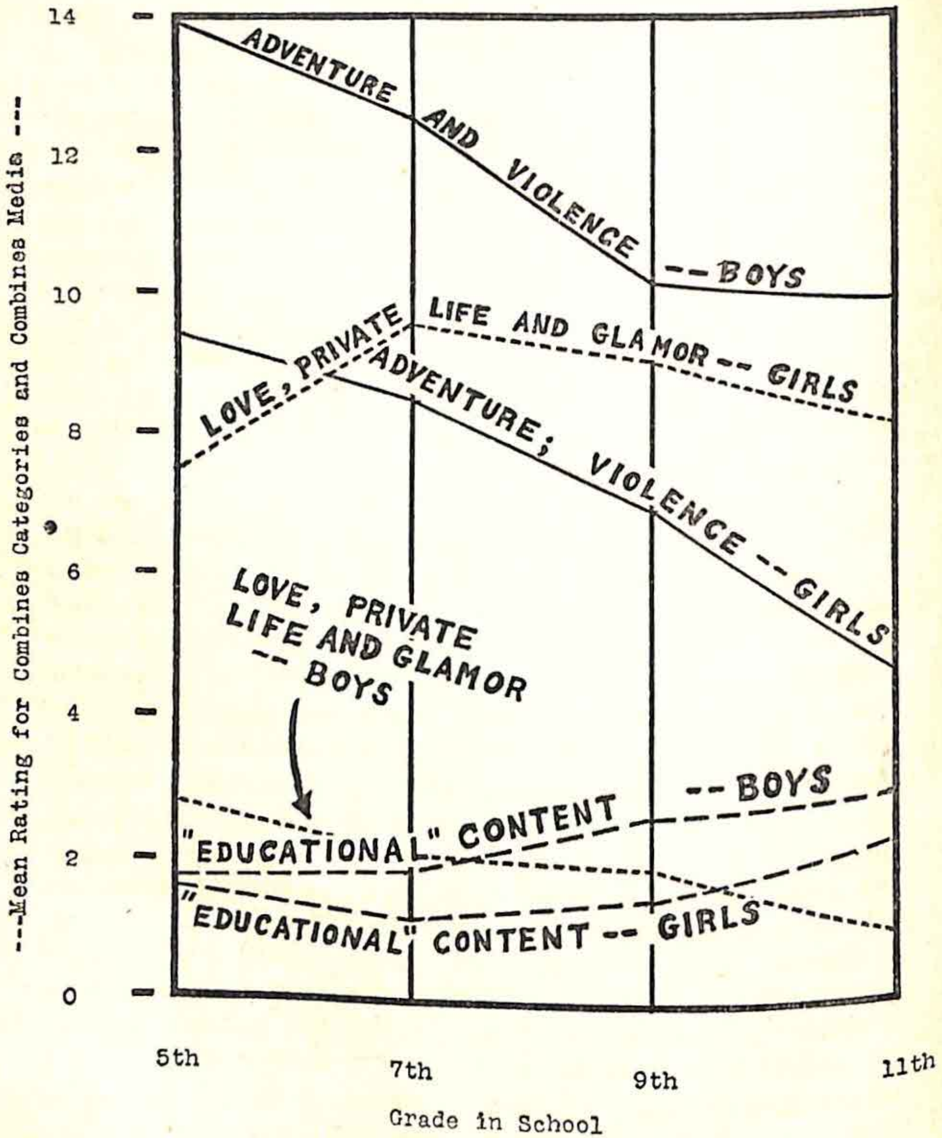


Figure I. Comparison of Three Hypothetical Patterns of Interest, by Preference Level, Sex and Grade.

making up each pattern occurred in more than one medium; for example, mystery occurred in radio, movies, magazines and books. Accordingly, the interests discussed in this section reflect subject-matter preferences in all media combined.

Examination of individual categories revealed certain distinct trends related to sex and grade differences. Within the adventure and violence pattern, interest in westerns dropped sharply with increasing age levels. Fifth-grade boys and girls gave westerns a far higher rating than any other category; yet eleventh-grade boys and girls gave westerns a relatively low rating. The same tendency could be observed in ratings of radio adventure serials. The ratings for mystery, crime, disaster, war, and adventure other than radio serials, remained at about the same level from grade to grade. Also, interest ratings for these five categories were roughly of the same magnitude within grades. These tendencies applied to both sexes, although in all cases boys gave higher ratings to adventure and violence than girls. Both sexes rated horror lowest among all categories classifiable as adventure and violence.

All individual categories classified under love, private life and glamor were given relatively inconsequential ratings by boys. However, some interest in movies and Hollywood was shown by boys, especially in the fifth and seventh grades. Girls gave high ratings to love and romance, and movies and Hollywood. A tendency for the love and romance ratings to rise as the girls' grade level advanced was noted Conversely, interest in movies and Hollywood appeared to drop progressively as the girls' grade level advanced. Considerably less female interest was shown in the remaining categories classified under love, private life and glamor; that is, radio love serials, news on love and marriage problems, newspaper columns on love and personal problems, radio columnists, and newspaper gossip columnists.

Eighteen individual categories were classified under 'educational' content. With a few exceptions, these categories were given low ratings by both sexes. Relatively high ratings were given science and international and foreign news by ninth- and eleventh-grade boys. History elicited a moderate preference rating from boys, regardless of grade, and from eleventh-grade girls. Ninth- and eleventh-grade girls also ranked international and foreign news notably higher than other 'educational' content. Fifth-grade girls showed a substantial preference for art, literature and music. These exceptions of course, were not reflected in the mean ratings shown in Figure I, because of the generally low ratings of all other categories under 'educational' content.

STRENGTH OF EACH PATTERN BY MEDIUM

An indication was desired of the consistency of interest in the three hypothetical patterns from medium to medium. Accordingly, the preference ratings of all categories applying to each pattern within each medium were averaged. For example: In order to obtain the adventure and violence rating for radio, the preferences for the three pertinent radio categories were averaged; that is, adventure serials, mystery plays, and western shows. For movies, there were five pertinent categories—adventure, horror, mystery, war, and westerns.

A considerable degree of consistency of interest from medium to medium was observed for certain grade-sex groups. Adventure and violence in all media appealed strongly to fifth- and seventh-grade boys. Love, private life and glamor in all media but radio appealed strongly to seventh- and ninth-grade girls. 'Educational' content in all media, except magazines, received a low preference rating among boys of all ages. The somewhat higher male interest in 'educational' magazines, was due to the substantial popularity of science and mechanics. The girls' ratings of 'educational' content were uniformly low in all media. On the other hand, marked media differences in the appeal of the hypothetical patterns to certain grade-sex groups were observed.

Analysis of differences in the adventure and violence classification revealed that the older the student the less consistency from medium to medium. Adventure and violence in radio and magazines was much less popular among eleventh-grade boys than fifth-grade boys, although male interest in movie and book adventure and violence continued at a high level in all grades. Girls' preferences for this pattern followed the same general tendency; that is, for media differences to develop in the upper grades. Little interest on the part of either sex was shown in adventure magazines. This may have been due to the relative vagueness of the category.

Girls' ratings of the love, private life and glamor categories revealed a tendency for magazine stories and features and movies to become more popular relative to the other media from the fifth through the eleventh grade. Conversely, the relative popularity, for girls, of magazines devoted entirely to love and glamor appeared to decrease sharply after reaching a peak in the seventh grade. Radio categories in the love, private life and glamor area had no significant appeal for either sex at any grade level.

Under love, private life and glamor, radio and books were the weakest media. The two radio categories were love serials and radio columnists. The book category was love and romance.

The male preference for "educational" content in magazines, as against other media, was due in large part to an interest in magazines on scientific and mechanical subjects.

THIRD GRADE PREFERENCES

Questions used in the third-grade interviews were largely of the free answer type; no check lists were used. Hence, the results were analyzed separately from those obtained in the higher grades.

Third-grade tastes, as revealed by this study, were focused mainly on the comics, westerns, animal themes, the world of make believe, adventure, comedy and humor, and mystery.

By medium, the prevailing choices of both boys and girls were comics in the newspaper; westerns on the radio and in the movies; fairy tales and fantasy in books; adventure serials on the radio; comedy and humor in movies and magazines; animal themes in magazines and books, and mystery in radio fare, movies and magazines.

Boys, more than girls, favored western movies and books; and mystery on the radio. Girls, more than boys, favored comedy and humor in movies and books; and animal themes in books.

Third-graders indicated little preference for love, private life and glamor, although a considerable number expressed their liking for the movie and radio sections of the newspaper. They also show little interest in informative, non-entertaining content in any medium.

AN ANALYSIS OF A PROCEDURE USED IN THE TEACHING AND RATING OF ORAL EXPRESSION

FRANCIS J. DI VESTA and JAMES H. L. ROACH

Air University
Maxwell Air Force Base, Alabama

The Air University's Air Command and Staff School, like other schools of higher learning, has been faced with the problem of evaluating various types of seminar performance not easily measured by objective tests. Among these is the problem of rating students' performance in oral presentation. Several methods have been tried for such ratings. New methods are being tried. The purpose of the present study is to evaluate the results of one general procedure used in teaching and rating oral expression at the Air Command and Staff School. This material is presented because the unique character of the course and rating system may be of benefit to those engaged in similar educational activity.

DESCRIPTION OF THE COURSE

The course in oral expression has been established to develop, in Air Force officers, the ability to organize speaking material and to present it logically and forcefully. This objective arises from the very practical needs of Air Force officers to speak before civilian organizations, to participate in Air Force conferences or joint conferences with representatives of the other services and the departments and agencies of the Federal or State governments, to brief their commanding officers, and to participate in staff meetings. Because the length of the course is only fifteen class hours, the instruction is oriented around formal oral presentation on the assumption that speaking before large groups would be the most difficult type of requirement to be met in the field. Only the elementary fundamentals of oral expression are taught. Students are not expected to be finished conference members or speakers at the end of the course.

The instruction includes three one-hour lectures on essential steps in preparing a speech, divisions and organization of a speech, and effective speaker projection. The remaining class periods are devoted to presentations by students. Each student is required to make four oral presentations. The first requirement is a five-minute talk on some phase of preparing a speech. The second requirement is a ten-minute talk on a subject of the student's choice, but of military significance. The third requirement is a five-minute impromptu presentation on the student's choice of one of four subjects within the broad areas of operations, personnel, logistics, and world events. This is a test of the student's ability to delimit a broad subject and to speak spontaneously on that subject. The fourth requirement is a prepared fifteen-minute talk wherein the student regards himself as a numbered Air Force staff officer required to give a speech on some phase of the military establishment to a civilian audience.

The total enrollment of four hundred sixty-five students is divided into small groups of about ten students each for the oral presentations. An instructor is assigned to each group. Both instructor and students rate each oral presentation on a rating scale described later in this report. All students are required to make the first three presentations. Those students in the top twenty per cent of the class are exempted from making the last presentation. Each rating provides a basis for grading the student and for constructive criticism.

DESCRIPTION OF THE RATING SCALE

The rating scale, devised by the Evaluation Branch of the Air University's Educational Advisory Staff, was made up of twelve items rated on a scale of one to five. Items were selected on the basis of characteristics which, it was believed, could be observed and remembered during the time the speaker was on the platform and which could be used as a guide for the speaker while preparing his speech. The number of items was limited to reduce rating time as much as possible and still provide sufficient coverage of qualities making for good performance.

These considerations were necessary since student-raters were being used and a number of ratings were made during each class period. The rating scale follows:

EVALUATION FORM FOR ORAL PRESENTATIONS

I. *Platform Presence.*—Does speaker appear poised and purposeful? Is he well groomed?

1 2 3 4 5

Ill-at-ease and untidy appearance detracts from platform presence.	Slightly self-conscious; presentable; acceptable platform presence.	Poised and immaculate; excellent platform presence.

II. *Introduction.*—Is introduction effective in getting attention, and audience acceptance? Is objective understood? Did he establish importance of subject and lead smoothly into it?

1 2 3 4 5

Inadequate introduction; failed to accomplish one or more of the purposes of an introduction.	Satisfactory introduction, fairly complete; adequate but not effective.	Complete and very effective introduction.

III. *Actions.*—Do posture, body movements, and gestures increase effectiveness of presentation?

1 2 3 4 5

Unnatural, awkward actions; annoying mannerisms definitely distract from presentation.	Fairly natural actions; annoying mannerisms slightly distracting.	Natural actions greatly aid presentation; no distracting mannerisms.

IV. *Eye Contact.*—Does speaker look at each member of the audience?

1 2 3 4 5

Eye contact detractive; tended not to look at audience.	Eye contact fairly effective; tended to look at audience	Eye contact very effective, well distributed, constantly looked at audience.

V. *Voice.*—Is voice pleasing? Can you easily and clearly hear him? Is he fluent and forceful?

1 2 3 4 5

Voice boring, irritating; indistinct, hesitant delivery.	Voice pleasing, fairly clear; matter of fact delivery.	Voice very pleasing and clear; eloquent, stimulating delivery.

VI. *Phraseology.*—Does he pronounce and use words correctly? Are sentences short, clear and complete? Does he use terms clear to audience?

1	2	3	4	5
Annoying grammatical errors. Poor choice of words. Does not define terms; difficult to follow.		Minor grammatical errors. Satisfactory choice of words; meaning generally clear.		Excellent grammar and sentence structure; effective choice of words; easily understood.

VII. *Body.*—Is plan of presentation clear to audience? Are major points effective and properly supported by minor points and details?

1	2	3	4	5
Plan not clear, too many or too few points inadequately supported.		Plan fairly clear; fairly adequate major points, moderately supported.		Plan very clear; proper number of major points, well supported.

VIII. *Transition.*—Does speech move smoothly and effectively from each part or point to following part or point? Does it have continuity?

1	2	3	4	5
Disjointed transition; parts not fitted together; lacks continuity.		Fair transition between parts; fair continuity.		Smooth transition; parts fit naturally together; excellent continuity.

IX. *Knowledge of Subject.*—Did speaker have sufficient background of facts on subject? Did he understand implications of material? Were his conclusions sound? Did he adequately cover and emphasize important aspects of subject?

1	2	3	4	5
Lacked essential knowledge of subject and its implication; conclusions faulty; failed to cover important aspects.		Fair knowledge of subject; conclusions generally adequate; partially covered important aspects.		Knew subject thoroughly; drew sound conclusions. Effectively covered important aspects.

X. *Conclusion.*—Does conclusion effectively end speech? Does it have punch? Is it consistent with purpose?

1	2	3	4	5
Weak, inadequate conclusion; new material introduced. Leaves audience up in air, not consistent with purpose.		Fairly effective ending, fairly consistent with purpose; lacks punch.		Concise, forceful ending; drives home purpose.

XI. *Preparation.*—Was speech extemporaneous (not memorized), and was it delivered without excessive use of props such as manuscript, notes or lectern? Was it within time limits and was proper time apportioned to developing each point or part? Were visual aids effectively used?

1	2	3	4	5
Indicated inadequate preparation by excessive use of props; poorly apportioned timing with speech too long or too short; faulty use of visual aids.		Indicated fair preparation by moderate use of props; fairly good timing; acceptable use of visual aids.		Indicated thorough preparation by no reliance on props; excellent timing and effective use of visual aids.

XII. *Interest.*—Did speaker establish and maintain audience interest in his subject and in himself as a speaker? Was audience reaction to speaker and subject friendly?

1	2	3	4	5
Dull and boring speaker, no audience interest; audience reaction indifferent or unfriendly.		Moderate audience interest; audience reaction fairly cordial.		Intense audience interest; audience reaction very friendly.

The grade for any one oral presentation was composed of the average of all students' ratings on this form added to the instructor's rating. Ratings are corrected for rater bias by simply adding or subtracting the rater's deviation from the mean of all ratings. Instructor ratings and students' ratings are corrected independently.

An indication of the reliability of ratings made on this scale was determined by finding the correlation between the average of student ratings and instructor ratings on each student's presentation in the second oral requirement (a ten-minute prepared speech on a subject of military significance). The correlation between four hundred sixty-five scores paired in this way was .63 ± .02. Although this correlation is not as high as is generally desired for reliability coefficients, it does represent a significant relationship of better than moderate magnitude.

Further information about the rating scale is presented in Table I where intercorrelations of ratings on each characteristic with ratings on every other characteristic and the total score are shown. The item numbers in this table correspond to the item numbers in the rating scale.

TABLE I.—INTERCORRELATIONS BETWEEN SCORES ON EACH OF THE TWELVE FACTORS OF THE EVALUATION FORM FOR ORAL PRESENTATION AND THE TOTAL SCORE*
(N = 465)

Item Number†	Rating Scale Item Number*												Total
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	
I		.46	.57	.47	.55	.48	.42	.61	.44	.44	.35	.53	.73
II			.33	.29	.37	.35	.44	.42	.42	.40	.29	.48	.62
III				.52	.42	.36	.30	.38	.41	.31	.26	.42	.62
IV					.38	.34	.11	.41	.35	.31	.26	.44	.61
V						.47	.36	.42	.38	.38	.33	.46	.65
VI							.42	.51	.47	.45	.37	.46	.68
VII								.68	.68	.50	.50	.63	.75
VIII									.68	.50	.54	.58	.79
IX										.50	.53	.62	.77
X											.56	.56	.70
XI												.46	.65
XII													.78

* Data is from the instructor ratings on Oral Presentation III only.
† Item numbers correspond to numbers opposite the rating scale items shown in the text.

These intercorrelations indicate the existence of a positive relationship among all items. Relatively higher relationships exist between each item score and the total score. These relationships are desirable because they indicate that, in general, each item evidently makes some unique contribution to the total score. None of the items correlates so highly with the total score that it could replace the entire scale. Items VII, VIII, IX, and XII have relatively higher intercorrelations than other items. This is to be expected, since they are also related in the practical situation. Each of these four factors is essential to preparation of the subject-matter content of the speech. These results have not been factorized, however, to determine the extent to which the assumed factors or other factors actually are present.

RESULTS

The effects of the oral presentation program are illustrated in the data obtained from the rating scales.

Table II shows the mean scores and standard deviation for each of the oral presentations. The scores shown are a composite of the average of nine student ratings plus one instructor rating. The total possible score is thereby 120. From this table it will be recognized that the rating scale reflected an increase in scores as the course progressed. Differences in mean scores between any one speech and each of the other speeches are significant at the one per cent level.

TABLE II.—MEANS AND STANDARD DEVIATIONS OF THE ORAL PRESENTATION SCORES*

Speech Number	Number	Mean	SD
I	460	85.8	10.75
II	460	90.9	9.45
III	460	89.1	9.82
IV	360	92.5	8.16

* Differences between means on Oral Presentations I and II, I and IV, II and III, and III and IV are significant.

The lower mean score for Speech III may be interpreted as evidence of the validity of the scale. Speech III is the five-minute impromptu speech with only ten-minute preparation. On this basis it would be expected that the individual's work would be poorer than on a prepared presentation. This interpretation is reflected in the lower mean score for the third speech than for the second or fourth speech.

Intercorrelations between each speech and every other speech are shown in Table III. These correlations show a substantial relationship between student performance early in the course and later performance. Lower correlations between scores on Speech IV and scores on other speeches are undoubtedly a result of the increased homogeneity of the class (the top twenty per cent on three earlier speeches were excused from making the fourth speech).

TABLE III.—INTERCORRELATIONS BETWEEN EACH OF THE ORAL PRESENTATIONS AND FINAL ORAL PRESENTATION SCORES
CLASS 49-B, AC&SS

Oral Presentation	Oral Presentation I N = 460	II N = 460	III N = 460	IV N = 360	Final Score N = 460
I		.67	.58	.40	.85
II			.69	.49	.89
III				.42	.84
IV					.71
Final Score					

A study was made of the progress of those individuals in the highest ten per cent of the class (N = 46) and those in the lowest ten per cent of the class (N = 46) selected on the first presentation. The averages for these groups are shown in Table IV. The significance of the differences between the scores for each speech for each of these groups is shown in Table V.

Although there is a drop of five points from average of Speech I to Speech III for the highest ten per cent, the difference is significant at slightly more than the five per cent level of con-

TABLE IV.—MEANS OF SCORES FOR EACH ORAL PRESENTATION MADE BY THE HIGHEST AND LOWEST 10 PER CENT OF STUDENTS ON ORAL PRESENTATION NO. I

Speech	Highest 10% on Speech I	Lowest 10% on Speech I
I	105.2	67.4
II	101.3	79.2
III	99.9	80.8
IV		86.7

fidence. None of the other differences in speech scores for this group is significant.

The lowest ten per cent of the class on Oral Presentation I made gains significant at least at the two per cent level of confidence except for the progress shown from Oral Presentation II to III. Again, this is probably accounted for by the fact that Oral Presentation III was an impromptu requirement.

TABLE V.—SIGNIFICANCE OF DIFFERENCES BETWEEN ORAL PRESENTATION AVERAGES MADE BY THE HIGHEST AND LOWEST 10 PER CENT OF THE STUDENTS SELECTED ON THE BASIS OF SPEECH I SCORES
(N = 46 for each group)

Differences Between Speeches:	Highest 10%			Lowest 10%		
	DF	X^2	P	DF	X^2	P
1 and 2	2	5.21	$>.05$	2	13.21	$<.01$
1 and 3	2	3.28	$<.20$	2	19.20	$<.001$
1 and 4				1	32.76	$<.001$
2 and 3	2	.07	.98	2	1.66	$<.50$
2 and 4				2	14.28	$<.001$
3 and 4				3	9.63	$>.02$

RELATIONSHIP BETWEEN PERFORMANCE IN ORAL EXPRESSION AND OTHER FACTORS

Correlations were found between the final oral expression grades and several other factors on which scores were available (see Table VI). The data studied were as follows: *General Survey Test*—a two-part test yielding a score on student achievement in Air Force background subject matter, another score for achievement in mathematics, science, and English and a total score; the *American Council on Education Psychological Examination's* Q, L, and total score; the student officer's rank; the student officer's period of active service; the student officer's age; consolidated scores of fifteen class examinations; and grades on the final school problem, which is a simulated problem situation solved by a staff of sixteen student officers.

TABLE VI.—CORRELATION OF FINAL ORAL EXPRESSION GRADES AND TWELVE FACTORS
(N = 465)

Factor	Correlation with Oral Expression Grades
General Survey Test	
Military subjects	.31
Basic	.26
Total	.33
ACE Psychological Examination	
Q score (quantitative)	.22
L score (language)	.30
Total	.29
Rank	.00
Months active service	.10
Age	.06
Class examinations (15 consolidated scores)	.47
Final school problem (staff problem)	.22
Final class standing (including oral expression)	.50

From these correlations, very little relationship is shown to exist between oral expression and rank, age, and months of active service. The oral expression grades have a higher relationship with general intelligence, as measured by the *ACE Psychological Examination* ($r = .30$), staff action ($r = .22$), and military achievement scores ($r = .31$). A still higher relationship is noted between oral expression and classroom achievement on written examinations ($r = .47$). It is interesting to note that the oral expression grades correlate more highly with the *American Council on Education Psychological Examination's* L score ($r = .30$) than with the Q score ($r = .22$). The difference between these correlations is, via Fisher's Z transformation, significant at the one per cent level of confidence.

SUMMARY

The oral expression course at the Air University is established for the purpose of imparting basic principles in the preparation and delivery of speeches. Practice is given through four speech requirements. Appraisal of student performance for purposes of evaluation and guidance is made by student and instructor ratings of each speech.

Results of the analysis of scores on the evaluation form for oral presentation indicate that (1) raters are differentiating between presentations; (2) growth and development in oral presentation takes place throughout the course; (3) no progress in oral presentation was reflected in the evaluation scale for those already qualified at this level of oral expression; (4) those who were originally very poorly qualified in oral expression developed to a point where their average score on the final speech requirement exceeded the entire class average on the first oral presentation. One might conclude that, as a group, the Air University students improve in oral expression throughout the period of instruction and, furthermore, that most improvement is made by those individuals who were poorest in oral expression before attendance at the Air University.

THE EFFECT OF ATTENTION-GAINING DEVICES ON FILM-MEDIATED LEARNING*

D. MORGAN NEU

The Pennsylvania State College

INTRODUCTION

Film-makers have employed a wide variety of techniques to achieve emphasis and increase the teaching effectiveness of their products. In general, the devices employed to direct attention towards particular points of the content presented may be divided into two broad classes: 'relevant devices' and 'irrelevant devices.' 'Relevant devices' include any emphasis technique that is related to, or operates upon, the specific point to be emphasized. For example, an ultra-close-up of part of a gun brings attention specifically and forcefully to that part of the gun. 'Irrelevant devices' include any emphasis techniques that call attention to the screen but are otherwise unrelated to the film content. For example, interrupting a film on a technical subject with a shot of a bathing beauty is an instance of an 'irrelevant device.'

While current learning theories and the experiments that support them are not entirely in accord on the relative effectiveness of these two methods of ensuring the perception and learning of the salient points in a lesson, the consensus seems to be that, in the long run, 'relevant devices' should facilitate recall, while 'irrelevant devices' should act as distractors and inhibit recall. This inference is obviously at variance with the results hoped for by producers who insert sequences that attract attention to the screen without relating to a specific point of information. It is important, therefore, to determine experimentally the effects on learning of such different attention gaining techniques.

STATEMENT OF THE PROBLEM AND HYPOTHESES

The major purpose of this study was to determine the relative effectiveness of three levels of attention-gaining devices; (1)

---

* Summary of a dissertation completed in pursuance of the requirements for the Ph.D. degree at The Pennsylvania State College, 1950. The work described here was accomplished under the Instructional Film Research Program under Contract N6 onr-269, Task order VII, with the Special Devices Center, Office of Naval Research.

relevant devices, (2) no devices at all, (3) irrelevant devices. Since it is possible to insert devices in the sound track or the visuals, the rôle of the medium of presentation has also been studied. In order to maintain a clear cut comparison between the sound and visual media, sound and visual attention devices were each studied separately. Three hypotheses were formulated for testing as follows:

1) *Device relevancy.*—The hypothesis is that film-mediated learning will be facilitated by the use of relevant attention-gaining filmic devices, and inhibited by irrelevant attention-gaining devices. The hypothesis was tested by comparing versions of a film containing, respectively, relevant devices, no devices, and irrelevant devices.

2) *Device medium.*—The hypothesis is that film-mediated informational learning will be equally facilitated by visual and sound devices of the same relevance. The hypothesis was tested by comparing learning from visual device versions with learning from sound device versions. This comparison is made separately for relevant and irrelevant devices.

3) *Device recall.*—The hypothesis is that recall of the devices themselves is relatively independent of learning of the factual information content of the film.

EXPERIMENTAL DESIGN AND PROCEDURES

The Experimental Films.—To accomplish the comparisons required by the hypothesis, five versions of a film presenting an introduction to machine shop measuring instruments were produced. The five versions included the following:

1) Basic (identified as B in the tables): contains no experimental emphasis devices, but gives a clear, straightforward treatment of the subject.

2) Visual relevant (VR): contains attention-gaining devices in the visuals that are related to the points of content being emphasized.

3) Visual irrelevant (VI): contains attention-gaining devices in the visuals that are unrelated to the points of content being emphasized.

4) Sound relevant (SR): contains attention-gaining devices in the sound track that are related to the points being emphasized.

5) Sound irrelevant (SI): contains attention-gaining devices in the sound track that are unrelated to the points being emphasized.

Choice, number, and placement of the devices.—To ensure satisfactory comparability among the four 'device' versions, a panel of judges selected four sets of devices considered to be comparable in intensity and probable effectiveness in gaining attention.

It was decided that the devices selected should not add information to the film, but should be used solely for their possible attention-gaining value. Furthermore, it was decided to select attention-gaining devices that would represent various categories of stimulus conditions: movement and change, size, intensity, repetitive or prolonged contact, novelty and peculiarity, impressiveness, and sudden contrast.

A device was inserted at each of the same twenty-six most important points in each version. Table 1 lists the devices and their placement in relation to the informational content of the film. Thus, a point emphasized in one version by a visual relevant device (e.g., a spotlight) is also emphasized in the other three versions, by appropriate devices. It will be noted that some devices are used more than once.

THE TESTS

Two tests were constructed to provide measures of (1) learning of information and (2) recall of devices. The information test comprised one hundred four multiple-choice items, of which twenty-five provided picture-choices rather than verbal choices. This test had a high reliability (.94 by Kuder-Richardson Formula 20).

The device recall test included seventy-seven items, each of which briefly identified a device by a short statement. The subjects were called upon to check whether (1) the item was in the film, (2) the item was not in the film, or (3) they did not remember whether it was in or not.

THE POPULATIONS

The experiment was conducted with two separate military populations: (1) 1576 Army recruits at Fort Dix, New Jersey, (2) 1055 Navy recruits at The Great Lakes Naval Training Station, Great Lakes, Illinois.

TABLE 1.—DESCRIPTION OF ATTENTION-GAINING DEVICES

Basic Version	Attention-gaining Devices Added to Basic Version			
	Visual Relevant	Visual Irrelevant	Sound Relevant	Sound Irrelevant
1. Three metal blocks, rectangular, square, and round, are used to illustrate length, width, and depth.	White line appears on each block to emphasize length, width, and depth.	Concentric circles flash on screen as length, width, and depth are mentioned.	Long, short, and low tones emphasize length, width, and depth.	Auto horn sounds as length, width, and depth are mentioned.
2. The scale graduations on each end and both sides of the steel rule are shown one at a time.	Each scale graduation is spot lighted as it appears.	Train-crossing signals flash on before each graduation appears.	A female voice names each graduation as it appears.	A train whistle sounds before each appearance of the graduations.
3. A hook rule is used to measure from the shoulder of a metal block.	The rule appears to move by itself onto the shoulder.	Still shot of an athlete in action appears before and after rule is put on shoulder.	Before hook rule appears commentator says, "Now hear this. You are to remember. . . ."	Commentator loudly clears throat before discussing hook rule.
4. A slender flexible rule is held up for close-up shot.	End of rule is bent back and released which causes it to vibrate.	Portrait of pretty girl appears before flexible rule is discussed.	A 2nd male voice gives the name of the rule.	Sound of buzzing airplane before flexible rule is described.
5. The steel rule is shown so that the worn corners on one end are plainly visible.	Extreme close-up of this worn end.	The entire scene is tinted pink.	The commentary is read, slowly with emphasis.	Sound of applause in the background of entire scene.
6. The steel rule is wiped with an oiled rag several times.	Picture of oil can is flashed on the screen with the mentioning of oil.	Picture of cow's tail.	A squeaking sound is heard as oiled rag is rubbed across rule.	Fire siren starts when rule is rubbed with oiled rag.

7. *The spring caliper is closed properly by squeezing the legs together.*	*Picture of hand squeezing the 1st and 2nd fingers of other hand together.*	*Concentric circles appear as caliper legs are closed.*	*Crunching sound as caliper legs are closed.*	*Sound of music as caliper legs are closed.*
8. A group of adjustable precision instruments (micrometer and vernier caliper) are shown on workbench.	As each is mentioned, one micrometer and one vernier caliper turns around by itself.	Picture of a hammer hitting a nail as each instrument is mentioned.	The word micrometer, said with a rising inflexion, is followed by an ascending whistle. After the word vernier caliper is said with falling inflexion, a descending whistle follows.	After each word is said, a gurgling sound-like pouring liquid from a jug is made.
9. Two parts of the micrometer, the barrel and thimble, are discussed.	Finger points to the thimble.	Shot of a smoke stack before thimble is mentioned.	A 3rd male voice names the thimble.	Sound of squeaking door before thimble is mentioned.
10. A piece of paper is drawn between the micrometer's measuring surfaces to clean them.	Close-up of paper being drawn through.	Picture of turkey's head appears.	Resinous twang is heard as paper is pulled through.	Sound of cuckoo clock as paper is pulled through.
11. Micrometer is adjusted to show that graduations should line up to zero when measuring surfaces are closed.	Spot-light on measuring surfaces and on scale graduations.	Shot of close-up of two eyes looking at camera.	2nd male voice reads commentary.	Sound of sawing wood.
12. Micrometer is twirled to show "wrong way" of obtaining rapid adjustment.	Micrometer is twirled very rapidly.	Picture of man getting hit in the jaw.	Ratchet sound while micrometer is twirled.	Sound of miscellaneous objects falling down and breaking. (Fibber McGee's closet.)

TABLE 1.—*Continued*

Basic Version	Attention-gaining Devices Added to Basic Version			
	Visual Relevant	Visual Irrelevant	Sound Relevant	Sound Irrelevant
13. Micrometer approaches moving object in lathe as if to measure it.	Restraining hand stops the micrometer.	Scene goes in and out of focus.	Commentator says, "Now watch this closely."	Sound of jazz music.
14. A micrometer which measures from three to four inches is called a four-inch micrometer.	Six-inch steel rule is placed along side micrometer.	Concentric half-circles appear on the screen.	Female voice reads commentary.	Sound of water splash and gurgling effect.
15. The vernier caliper is used to measure a part in a lathe.	Part of scene at unusual angle from under the operator's arm.	Picture of profile of man' face.	Background sound of lathe running.	Sound of baby crying.
16. Snap gage is shown in comparison with a micrometer in background.	Snap gage is moved to extreme close-up in foreground with micrometer in background.	Pitcher throws ball at camera.	3rd male voice reads commentary.	Sound of breaking glass.
17. The three part of the snap gage (anvil, go button and not-go button are described as gage is shown on workbench.	Finger points to each part as it is mentioned.	Concentric half circles appear before each part is mentioned.	Sound of anvil when anvil is mentioned, sound of "go" police whistle for go button, and "stop" whistle for not-go button.	Sound of horse whinnying as each part is mentioned.

18. A part is inserted in a ring gage several times.	Zoom to close-up.	Picture of man lying on ground, feet nearest camera.	Ascending whistle when part is inserted in gage.	Pistol shot when part is inserted in gage.
19. Snap gage is inspected for accuracy with gage blocks.	Circular mask to concentrate attention on gage blocks.	Picture of locomotive coming at camera.	2nd male voice reads commentary.	Sound of fog horn.
20. Straight plug gages and taper plug gages are shown on workbench.	All but one straight gage disappear and reappear. All but one taper gage disappear and reappear.	Picture of baby as each group of gages is mentioned.	After "straight plug gages" said with rising inflexion, an ascending whistle. After "taper plug gages" a descending whistle.	Sound of drum roll as each group is mentioned.
21. Shot of hole in metal part. Hole contains many burrs.	Burrs gradually disappear by a dissolve.	Picture of airplane diving through sky.	Sound of filing metal as burrs disappear.	Sound of telephone as burrs disappear.
22. Measuring instruments are being checked in inspection laboratory.	Shot of large white check mark.	Picture of hammer hitting nail.	Voice says, "O. K."	Sound of hammer hitting nail.
23. Gage blocks are used to check setting of fixed gage.	Unusual angles: close up of gage with inspector's eye appearing through curve of gage; gage is seen in silhouette.	Still picture of athlete in action.	Female commentary.	Sound of rooster crowing.
24. Two gage blocks are "wrung" together.	Circular mask makes operation appear as through a hole.	Picture of gun being fired at camera.	Sound of woman's scream as blocks are "wrung."	Sound of coughing as blocks are "wrung."
25. Shot of projection comparator.	Hand points to the comparator.	Picture of girl's legs.	2nd male voice reads commentary.	Sound of "gibberish" (Donald Duck).
26. Shot of the interference bands produced by the light wave instrument.	Extreme close-up of the interference bands.	Picture of boy and girl kissing.	Sound of voice counting "1, 2, 3, 4, 5."	Sound of kiss and sigh.

PROCEDURES

At each testing center, the following procedures were employed: intact companies of recruits were divided at random into six comparable groups. Five of these groups were each shown one of the film versions (B, VR, VI, SR, or SI) and, immediately after the film, given the information tests and the device recall test. The sixth (control) group was given the information test without having seen the film. Test responses were recorded on IBM answer sheets, and machine-scored.

STATISTICAL TREATMENT OF THE DATA

For the men in each population both a measure of general intelligence and a measure of mechanical aptitude were available. Mean scores for the information test for the six treatments were calculated, and these means were adjusted, by an analysis of co-variance, to take into account small differences in initial ability as measured by the general intelligence and mechanical aptitude tests. The adjusted means therefore reflect the performance to be expected of groups having equal mean general intelligence scores and equal mean mechanical aptitude scores.

To determine the relationship between the information test and the device test, scores on these two tests were correlated.

RESULTS

Comparison of the Film Versions

Information Test Results.—In Table 2 the mean scores on the information test are given for the six groups of enlisted men in the Army and in the Navy. The adjusted means are given in parentheses. Table 3 summarizes the differences among the adjusted information test mean scores.

The findings can be summarized for both the Army and Navy populations as follows:

1) For all film groups substantial learning resulted from viewing the film. In each case, the film group mean score was between one and two standard deviations higher than the mean score of the comparable control group that saw no film.

2) The 'sound irrelevant' group had a lower mean score, in each population, than any other film group. It is to be noted, however, that the visual irrelevant version was the most effective

TABLE 2.—RAW MEAN SCORES AND ADJUSTED MEANS OF NAVY AND ARMY POPULATIONS ON INFORMATION TEST AFTER VARIOUS FILM TREATMENTS

Treatment	Navy			Army		
	N	Mean	SD	N	Mean	SD
Visual Relevant (VR)	136	67.63 (69.55)*	15.18	248	47.84 (48.02)	17.36
Visual Irrelevant (VI)	148	65.45 (66.62)	16.67	229	49.83 (49.24)	17.26
Sound Relevant (SR)	135	70.64 (70.68)	14.19	226	48.12 (49.00)	16.31
Sound Irrelevant (SI)	155	66.67 (65.11)	15.17	242	46.35 (46.31)	17.71
Basic (B)	331	69.80 (69.52)	15.87	367	52.55 (52.09)	18.80
Control (C)	150	40.57 (39.88)	10.40	264	30.04 (30.41)	9.74

* Adjusted means are in parentheses.

TABLE 3.—DIFFERENCES AMONG ADJUSTED MEAN SCORES FOR INFORMATION TEST[a]

	VR	VI	SR	SI	B
			Navy		
VI	−2.93*	—	—	—	—
SR	+1.13	+4.06**	—	—	—
SI	−4.44***	−1.51	−5.57***	—	—
B	−.03	+2.90*	−1.16	+4.41***	—
C	−29.67***	−26.74***	−30.80***	−25.23***	−29.64***
			Army		
VI	+1.22	—	—	—	—
SR	+.98	−.24	—	—	—
SI	−1.71	−2.93*	−2.69*	—	—
B	+4.07***	+2.85**	+3.09**	+5.77***	—
C	−17.61***	−18.83***	−18.59***	−15.90***	−21.68***

[a] In this table, the mean for the group listed on the top is subtracted from that of the group listed on the left.

* Significant at the 5 per cent level of confidence.

** Significant at the 1 per cent level of confidence.

*** Significant at the 0.1 per cent level of confidence.

of all the versions containing attention devices for the Army population.

3) The remainder of the inter-film differences were inconsistent for the two populations, and generally insignificant. In the Army, the basic version was consistently more effective than any device version, while in the Navy it was about equal to the two relevant device versions.

4) No significant difference was noted as between the visual devices and the sound devices versions, combined without respect to relevance of devices. For the Navy, the weighted mean of the two visual devices versions was 68.02, as compared with a weighted mean of 67.70 for the sound devices versions. For the Army, the visual devices group mean, without respect to relevance of device, was 48.60, while the sound group mean was 47.61. (These are all adjusted means.)

Results of the Device Recognition Test

The recognition test items were divided into two groups for each film version: (1) those devices that were in the film (Test C_1) and (2) those devices that were not in the film (Test C_2).

In scoring responses for the C_1 test, which comprised the devices actually in the film, a plus one (+1) was given for each device reported as present, a zero (0) if the subject reported he did not remember, and a minus one (−1) if the device was reported not present. The reverse of the scoring of the C_1 test was used for the C_2 test which comprised all devices not in the version being tested. A plus one (+1) was given for each device reported as not present, a zero (0) if the subject did not remember, and a minus one (−1) for each device reported as present. These scores were totaled for each film version and the tests (C_1 and C_2) correlated with each other as well as with the film information test scores. Table 4 reports these correlations.

All the correlations between the film information test and the C_1 test, and the film information test and the C_2 test, are generally insignificant although positive, indicating that there was only a slight relationship between scores on the film test and the remembering of the devices. The correlation between the C_1 and C_2 tests is also low. In all cases except one, which was practically zero, the correlations were negative. Analysis of the data showed that there was a slight tendency for the individual to mark the 'not-present' devices as present in the

TABLE 4.—CORRELATIONS BETWEEN DEVICE RECOGNITION TESTS AND INFORMATION TEST: C_1 TEST-DEVICES IN THE VERSION, C_2 TEST-DEVICES NOT IN THE VERSION

Group	C_1 Test and Information Test		C_2 Test and Information Test		C_1 Test and C_2 Test	
	Navy	Army	Navy	Army	Navy	Army
Visual Relevant	.185	.191	.191	.330	−.295	−.079
Visual Irrelevant	.338	.299	.132	.344	.042	.053
Sound Relevant	.223	.092	.141	.203	−.431	−.225
Sound Irrelevant	.107	.075	.073	.150	−.416	−.282
Basic*			.214	.334		

* Since the Basic Version contained no devices there was no C_1 test score for this group.

film, if he marked the 'present' devices as present. In other words, there was a general tendency to indicate that devices were present, whether they were used or not.

An analysis of responses to the individual items included in the Device Recognition Test revealed two significant points. First, whereas only about half the items included in the VR, SR, and SI versions were recognized by seventy per cent or more of the groups seeing these versions, almost three quarters of the devices in the VI version were correctly recognized by seventy per cent or more of the groups seeing this version. In short, the visual irrelevant devices seemed to have called attention 'to themselves' to a greater extent than any other type of device did. Second, whereas a maximum of only seven devices were falsely recognized (i.e., they were not in the version but the subjects said they were) by forty per cent or more of the VR, VI, and SR version groups, fifteen items were falsely recognized by at least this proportion of the sound irrelevant group. It is suggested that the irrelevant sound devices tended to confuse the subjects so that they were uncertain about what they did hear.

SUMMARY OF THE RESULTS

The results may be summarized with respect to the three experimental hypotheses as follows:

Device relevancy.—There is no evidence that the insertion of relevant devices, as defined in this study, adds to the effectiveness of an informational film. There is some evidence that irrelevant sound devices detract from the teaching effectiveness of such a film. (For both populations, the sound irrelevant version yielded the lowest scores.) To a lesser degree there is evidence that any of the devices used may have the same effect. (For the Army populations the basic version—no devices—yielded the highest score.)

Device medium.—No significant difference was found between the effectiveness (or ineffectiveness) of visual and sound attention-gaining devices.

Device recognition.—Ability to recognize and remember which devices were used in a film version was practically independent of learning from the version.

CONCLUSIONS AND RECOMMENDATIONS

This study leads to the general conclusion that attention-gaining devices of the kinds inserted in these films do not add significantly to learning.

The following recommendations are suggested as a result of this investigation:

1) Where instruction is the principal aim, and cost a consideration, producers of training films should present the subject matter in a simple, straightforward way and avoid the use of such fancy and expensive devices as spotlighting, zooms, extreme magnification, and stop motion, to gain the learners' attention.

2) If it seems necessary, in an instructional film, to use devices to attract or direct the learners' attention, use a technique which will emphasize something already in the film—some special treatment of indigenous materials related to the subject content—rather than introduce extraneous or irrelevant materials.

The findings of this experiment are largely negative; however, that does not nullify the usefulness of the study. It is almost as important to discover that certain techniques are ineffective, as it is to learn that other techniques are effective. In general, this study corroborates the belief which many educators have held for a long time, that it is not necessary to have a fancy, expensive treatment to convey ideas by means of film, and that a simple straightforward treatment is often the most effective.

A COMPARISON OF READING ABILITY AND READABILITY

JOHN U. MICHAELIS and FRED T. TYLER

School of Education
University of California, Berkeley

Numerous attempts have been made to develop a satisfactory technique for determining the level of reading difficulty of printed materials. There are three of these which are given most extensive use at the present time; namely, those proposed by Lorge, Flesch, and Dale and Chall. Since their appearance many types of materials have been appraised by one or more of the formulae: textbooks and children's books;[20] advertising;[1] industrial publications;[13] pamphlets on international relationships;[10] newspaper materials;[18] corporate annual reports.[21] The formulae have been used by the American Library Association[4] and the United States Department of Agriculture Extension Service.[22]

Since these techniques essentially provide a type of 'measuring' instrument, they should be examined to determine their validity and reliability. The first study on reliability, i.e., 'analyst-to-analyst reliability, of the Flesch formula was reported by Haynes, Jenkins and Walker.[12] They conclude that "For practical purposes the Flesch formulas and the directions for their use are sufficiently objective to be used even by inexperienced analysts to obtain estimates of reading ease and human interest of written materials." As for validity, Flesch found that materials rated as more readable by his formula were also more readable according to readership surveys and to expert opinion. The studies which the present writers have located have relied upon this latter type of criterion of readability. The subjective nature of such a criterion raises the question of its reliability.

PURPOSE

For the present paper it is assumed that materials rated as readable at a certain grade level should be understandable to students at that level. Thus, if a readability formula indicates that certain material is located at the Grade IX level, then it is to be expected that children at the Grade IX reading level should be able to read the material and then to answer questions on it with a high degree of accuracy.

The purpose of the present paper is twofold: (a) to determine the readability of certain United Nations publications by each of the three readability formulae mentioned above; (b) to determine the reading ability of a group of Grade XII students by means of some standardized reading test, and the accuracy with which these same students can answer questions based on the United Nations materials.

The need for such a study has been emphasized in numerous papers appearing during the last few years. Lostutter[18] concluded that the Flesch formula was best adapted for the measurement of the readability of newsprint. Lorge[17] has stated that each of the three formulae has a specific function to perform. He claims that his formula will give "evidence on the comprehension difficulty of the kind of reading pupils have to do in school," that the Flesch formula provides a measure of readability of magazine materials, and that the Dale-Chall formula "will make for adequate appraisal of health materials." Flesch, on the other hand, has pointed out that, "There is no theoretical justification for this neat pigeonholing of formulas; all three of those mentioned were based on exactly the same criterion; namely, the McCall-Crabs' Standard Test Lessons in Reading." Both Lorge and Flesch agree, however, that there is some degree of correlation between the formulae when applied to the same materials: "The statistical differences between their predictive values are insignificant." In a recent paper Michaelis[19] indicated that some U. N. materials were written on a difficult level according to teachers' judgments and data from readability formulae. He reported also that there was considerable disparity between the difficulty levels obtained by the different formulae.

TEST MATERIALS

The Iowa Silent Reading Test (Form Cm) was administered to forty-two boys and thirty-six girls in Grade XII at the Oakland Technical High School to determine the grade level at which these students were reading. The same students were also given an objective test based upon several U. N. publications which had been prepared for use in high-school social studies classes.

The U. N. test included four subtests: vocabulary, sentence comprehension, paragraph comprehension, and reading for details. The first part required the students to read a paragraph chosen from one of the U. N. pamphlets. The students were allowed three minutes to read and study the materials. At the end of that time the printed materials were collected and the students attempted to answer questions based on their reading. There were nine multiple-choice questions in Part I. The second part was a vocabulary test, again multiple-choice in nature. The test included twenty-two words and phrases, such as "preamble," "scope," "indigenous," "the framers of the Constitusion," and "ocean rendezvous." The third part consisted of a series of sentences, a phrase in each sentence being underlined. The students were required to indicate which of four choices could most reasonably be substituted for the underlined expression. The last part attempted to measure the students' understanding of short paragraphs. The multiple-choice questions were arranged to test the ability of the student to select the best title for the paragraph; to test his knowledge of word meanings; to test his ability to read for details. There was also a free expression question on each paragraph, but this item is not used in the present report. The reliability of the test of sixty-nine items, determined by the split-half method and stepped-up by the Spearman-Brown Prophecy formula, was .85.

FINDINGS

The readability of the U. N. materials was determined by each of the formula listed above. The Lorge formula considers average sentence length, ratio of prepositional phrases to number of words, and ratio of hard words to number of words in various samples of the materials being appraised. The Flesch formula for reading ease involves the determination of number of syllables per one hundred words and average sentence length. The Dale-Chall formula requires the calculation of average sentence length and the number of hard words. Readability by each formula was determined on the same sampling of U. N. publications. Five selections were chosen from each of five pamphlets according to the directions for the application of the formulae. Similarly the proper number of words was included in each sample. The computation of readability was done on two separate occasions with practically identical results. Furthermore, five other samples were chosen from each of the five pamphlets. Readability was determined for each of these samples, the results checking with those obtained from the previous sampling. Also the U. N. test was prepared in the main from the materials that had been used to determine readability. The readability of the test materials and of the instructions for the test was also determined by each formula. The U. N. publications varied from seven to two hundred one pages in length and were believed to be representative of those available for use in the public schools.

TABLE I.—READABILITY OF UNITED NATIONS MATERIALS APPRAISED BY THREE FORMULAE

United Nations Material	Ratings According to Formulae					
	Dale-Chall		Flesch		Lorge	
	Raw Score	Grade Placement	Raw Score	Level of Difficulty	Raw Score	Grade Placement
Basic Facts	9.6226	13–15	28.6720	College graduate	8.6524	8+
Everyman's UNO	9.4727	13–15	24.2233	College graduate	8.3557	8+
FAO	8.2895	11–12	38.9930	13–16	7.5497	7+
Guide to UN Charter	9.8611	13–15	31.7220	13–16	8.1380	8+
How Peoples Work Together	9.5917	13–15	37.6400	13–16	7.9803	8
Average	9.3675	13–15	32.2500	13–16	8.1352	8+

Table I shows the readability of the five pamphlets computed by each of the three formulae discussed above. From the table it may be seen that the Lorge formula indicates that the materials should be readable at about the Grade VIII level. According to the Flesch formula the materials are rated at the college undergraduate and graduate levels. Readability as estimated by the Dale-Chall method is placed at Grades XI to XV for the different pamphlets. There is, then, considerable discrepancy between the results obtained from different readability formulae.

As has been mentioned, it has been assumed for this study that if materials are readable at a certain level, then students at that level of reading ability should be able to read the materials with a high degree of accuracy. Specifically this would mean that according to the Lorge formula children who read at about the Grade VIII level should be able to comprehend the U. N. materials. Table II shows the mean percentile rank on the Iowa Silent Reading Test of the seventy-eight subjects in Grade XII according to the manual of directions.

TABLE II.—MEAN STANDARD SCORES AND PERCENTILE RANKS OF SUBJECTS

Subtest	Mean Standard Scores	Percentile Ranks in Grade XII
1. Rate	171	41
1. Comprehension	203	92
2. Directed Reading	166	42
3. Poetry Comprehension	172	50
4. Word Meaning	179	56
5. Sentence Meaning	171	35
6. Paragraph Comprehension	176	58
7. Use of Index	177	63
7. Selection of Key Words	169	43

The mean standard score on the whole test was 174, which corresponds to a grade equivalent of 13.0, indicating that these students were on the average reading at about the Grade XIII level.

Before indicating the success of these subjects on the U. N. reading test, it is desirable to indicate the readability level of the test materials and of the instructions to the test. This is done in Table III.

The sampling of test materials was no more difficult than was the sampling of the U. N. publications; the instructions were readable at a slightly lower level than were the test materials. The U. N. test was slightly easier than were the pamphlets upon which it had been based because the readability score for the test included the estimate of the readability of the instructions. It now remains to indicate how well the students were

TABLE III.—READABILITY OF U. N. TEST AND OF TEST INSTRUCTIONS

U. N. Test			Test Instructions	
Formula	Raw Score	Interpretation	Raw Score	Inter-pretation
Dale-Chall	8.8625	Grades 11–12	6.7622	Grades 7–8
Flesch	45.3711	Grades 13–16	64.8980	Grades 8–9
Lorge	7.7462	Grades 7+	6.3335	Grades 6+

TABLE IV.—MEAN SCORES ON U. N. READING TEST

Subtests	No. of Items	Mean	Mean Percentage Correct
1	9	6.44	71.56
2	22	15.47	70.32
3	6	3.79	63.17
4	32	17.36	54.25
Total	69	43.06	62.41

able to perform on the U. N. reading test. Mean scores on each of the subtests are shown in Table IV.

In the absence of any absolute standard of how well students reading at the Grade XIII level should be able to perform on the U. N. reading test, it still appears reasonable to conclude that even at this level of ability the test offers considerable difficulty. It would seem that students who are reading below the Grade XIII level would experience even more trouble. Judged by this criterion of reading ability, the readability formulae appear to place the U. N. materials at too easy a level, some formulae showing a greater disparity between reading ability and readability than do others.

SUMMARY AND CONCLUSIONS

The present findings offer support to Lorge's contention that different formulae may be more suitable for some materials than for others. If this is so, there is need for research to determine the most appropriate formula to use with specific types of materials. There seems no doubt that factors other than those now being considered must be employed if suitable formulae are to be developed. As Lorge has pointed out, present techniques really do not assess the difficulty of the concepts found in the materials. Length of word and number of syllables are not necessarily correlated with abstractness of a concept, for some very abstract concepts involve few letters, whereas some concrete concepts are expressed by lengthy words.

The following conclusions appear to be justified by the evidence here presented:

1) There is considerable disparity between the readability levels found for U. N. materials by different readability formulae.

2) The Flesch and the Dale-Chall formulae estimate the readability of this type of material somewhat more accurately than does the Lorge formula, if level of reading ability is used as the criterion.

3) The U. N. materials sampled appear to be too difficult for high-school students.

Finally, it would appear that there is need for extended study of the readability of various types of materials intended for use in the public schools. These studies might well be carried out on regular types of school materials as well as on special materials such as pamphlets at various grade levels. Further study of the problem of readability is indicated, probably especially designed to consider the nature of the concepts involved. Lastly, it may be that the basic formulae are logically sound, and that the problem is the development of more valid norms.

BIBLIOGRAPHY

1) Julian Alden. "Lots of Names—Short Sentences—Simple Words," *Printers Ink*, 1945, 29.

2) Edgar Dale and J. S. Chall. "Formula for predicting readability: instructions," *Educational Research Bulletin*, 1948, 27, 11–20, 37–54.

3) J. N. Farr and J. J. Jenkins. "Tables for use with the Flesch Readability Formula," *Journal of Applied Psychology*, 1949, 33, 275–8.

4) Pauline J. Fihe, Viola Wallace and Martha Schulz, Compilers. *Books for Adult Beginners; Grades I to VII*. (Rev. Ed.) Chicago: American Library Association, 1946.

5) Rudolf Franz Flesch. "A dissenting opinion on readability," *Elementary English*, 1949, 26, 332–4.

6) ———. "A new readability yardstick," *Journal of Applied Psychology*, 1948, 32, 221–233.

7) ———. "Readability formula in practice," *Elementary English*, 1945, 25, 344–51.

8) ———. *The Art of Readable Writing*. New York: Harper and Brothers, 1949.

9) Alberta S. Gilinsky. "How Valid is the Flesch Readability Formula?" *American Psychologist*, 1948, 3, 261. (Abst.)

10) S. N. Gluckenheimer. "Readability of pamphlets on international relationships," *Educational Research Bulletin*, 1947, 26, 231–8.

11) Robert Gunning. "Gunning finds papers too hard to read," *Editor and Publisher*, 1945, May 19, 12.

12) P. M. Hayes, J. J. Jenkins and B. J. Walker. "Reliability of the Flesch Readability Formulas," *Journal of Applied Psychology*, 1950, 34, 22–26.

13) J. J. Jenkins. "Readability of written communications in industrial situations," *American Psychologist*, 1948, 3, 356. (Abst.)

14) Bryant Kearl. "A closer look at readability formulas," *Journalism Quarterly*, 1948, 25, 344–348.

15) Irving D. Lorge. "Lorge and Flesch Readability Formulae: Correction," *School and Society*, 1948, 67, 141–2.

16) ———. "Predicting readability," *Teachers College Record*, 1943–44, 45, 404–19.

17) ———. "Readability formulae: evaluation," *Elementary English*, 1949, 26, 86–95.

18) Melvin Lostutter. "Some critical factors of newspaper readability," *Journalism Quarterly*, 1947, 24, 307–314.

19) John U. Michaelis. "United Nations Reading Materials," *Social Education*, 1950, 14, 221–222.

20) Leo R. Miller. "Reading grade placement of the first twenty-three books awarded the John Newbery Prize," *Elementary School Journal*, 1946, 46, 394–399.

21) S. Pasholian and W. J. E. Crissy. "How readable are corporate annual reports?" *Journal of Applied Psychology*, 1950, 34, 244–248.

22) U. S. Department of Agriculture, Soil Conservation Service. *Our American Land; the story of its abuse and conservation.* (Misc. Publ. No. 596.) Washington, D. C.: U. S. Government Printing Office, 1946.

NOTES ON A SUGGESTED INDEX OF ITEM VALIDITY: THE U-L INDEX

A. PEMBERTON JOHNSON

Educational Testing Service
Princeton, New Jersey

In 1935 Long, Sandiford, et al,* reported an empirical comparative study of the effectiveness of thirteen validity techniques in selecting from among four hundred eighty-two six-choice vocabulary items that set of fifty and that set of one hundred items which correlated highest with criterion scores based on all four hundred eighty-two items. For this situation they concluded, among other things, that those validity techniques which tend to discriminate in favor of items of fifty per cent difficulty appear to be superior to the techniques such as bi-serial r, Kelley and Vincent which make no discrimination in favor of median difficulty items. Considering both effectiveness and ease of computation, they concluded that among those techniques tending to select items of fifty per cent difficulty (percentage passing), the upper vs. lower (group) technique was to be preferred. They stated that item validities derived by the upper vs. lower technique tend to reach maximum reliability when the proportion of cases in each of these groups is twenty-seven per cent.† Numerically the item validities were expressed in this technique by the difference between number of correct responses for the highest twenty-seven per cent (or by the highest third) and the number of correct responses for the lowest twenty-seven per cent (or the lowest third) of the cases as arrayed on the criterion scale.

This paper proposes the U-L Index, a simple modification of the upper-lower twenty-seven per cent technique. This index proved extremely useful during World War II as a basis for selecting items to be included in Army Air Forces bombardier knowledge tests. The U-L Index effectively tends to select items of moderate difficulty which show good separation between upper and lower twenty-seven per cent groups. It is more easily computed and interpreted than any conventional item validity index known to the author. For many practical test construction situations where expensive item analysis cannot be justified, it is believed to be quite useful.

---

* John A. Long, Peter Sandiford, and Others. *The Validation of Test Items.* Bulletin No. 3, Dept. of Ed. Res., University of Toronto, Toronto, 1935, pp. 94-5.

† See also T. L. Kelley, "The Selection of Upper and Lower Groups for the Validation of Test Items," *J. Educ. Psychol.*, xxx, January, 1939, pp. 17–24.

This modification possibly has been used by other investigators before, but is believed not to have been. It requires the difference between the groups in number of correct responses to be divided by the number of cases in each group. Symbolically, if

R_u = number correct responses in the upper group of twenty-seven per cent
R_l = number correct responses in the lower group of twenty-seven per cent
N = number of cases
$f = .27N$

$$\text{U-L Index} = \frac{R_u - R_l}{.27N} = \frac{R_u - R_l}{f} \tag{1}$$

This index can theoretically vary from +1 through 0 to −1. These proportions are independent of the N of the population sample and similar in magnitude though not strictly in meaning to conventional correlation coefficients.

Long, Sandiford, et al., found in their empirical study that (1) the average difficulty of a group of items shown to be least valid by the bi-serial r technique was about equal to the average difficulty of the group of items indicated to be least valid by the upper-lower thirds technique, but (2) that the bi-serial r tended to eliminate items indiscriminately throughout the entire range of item difficulty whereas the upper-lower thirds technique tended to eliminate items farthest removed from the fifty per cent level of difficulty (percentage passing).*

In a table of intercorrelations among the validity values obtained for one hundred items by the thirteen methods which they studied, Long, Sandiford, et al., report a correlation of .65 between bi-serial r and upper-lower thirds validities. This finding strongly suggests the need to interpret the proposed

---

* Ibid., p. 114.

U-L Index in somewhat different fashion than the conventional bi-serial r.

The standard error of the U-L Index may be derived from the conventional formula for the standard error of the difference of two proportions, when the data are uncorrelated.

Guilford* gives the following generalized formulae:

$$\#34 \qquad \sigma_{d_p} = \sqrt{\sigma_{p_1}{}^2 + \sigma_{p_2}{}^2 - 2r\sigma_{p_1}\sigma_{p_2}}$$

$$\#28 \qquad \sigma_p = \sqrt{\frac{pq}{N}}$$

where: σ_{d_p} = S.E. of a difference between two proportions
σ_p = S.E. of a proportion
p_1 = 1st proportion
p_2 = 2nd proportion
$q = 1 - p$
N = number of cases
r = coefficient of intercorrelation of the two sets of data.

Using the previously mentioned symbols for the U-L Index:

$$\sigma_{p_1} = \sqrt{\frac{p_1q_1}{N}} = \sqrt{\frac{\left(\frac{R_u}{f}\right)\left(1 - \frac{R_u}{f}\right)}{f}} = \frac{1}{f}\sqrt{\frac{R_u(f - R_u)}{f}}$$
$$= \frac{1}{f}\sqrt{R_u - \frac{R_u{}^2}{f}}$$

Similarly
$$\sigma_{p_2} = \frac{1}{f}\sqrt{R_l - \frac{R_l{}^2}{f}}$$

In Guilford's formula #34 for σ_{dp}, the most conservative estimate of r is 0. That formula then reduces to $\sigma_{dp} = \sqrt{\sigma_{p1}{}^2 + \sigma_{p2}{}^2}$

By substitution:

$$\text{SE U-L Index} = \frac{1}{f}\sqrt{R_u - \frac{R_u{}^2}{f} + R_l - \frac{R_l{}^2}{f}}$$
$$= \frac{1}{f}\sqrt{R_u + R_l - \left(\frac{R_u{}^2 + R_l{}^2}{f}\right)}$$

The formula for the standard error of the U-L Index is such that the maximum value of the standard error occurs in the

---

* J. P. Guilford. *Fundamental Statistics in Psychology and Education*, New York: McGraw-Hill, 1942, pp. 133 and 142.

TABLE I.—5% AND 1% "t" TEST SIGNIFICANCE LEVELS FOR THE SE OF U-L INDEX WHEN THE U-L INDEX = 0 AND 50% PASS THE ITEM*

N	$f = .27N$	SE of the U-L Index	5% Level	1% Level
100	27	.136	.273	.364
148	40	.112	.223	.295
200	54	.096	.190	.252
248	67	.087	.172	.227
300	81	.079	.156	.206
348	94	.073	.144	.190
400	108	.068	.134	.177
500	135	.061	.120	.159
1000	270	.043	.084	.111

* For example, only 1 per cent of obtained values would be expected to exceed ±.364 were the true index zero for a sample of one hundred cases. In this table the degrees of freedom have been taken as equal to $2(f\text{-}1)$ since f is defined as the number in each of the two contrasted groups.

fifty per cent passing situation where the true index equals zero. Table I below gives the 5% and 1% "t" significance levels of the SE of the U-L Index in that situation.

A quite serviceable index of item difficulty (percentage passing) for each item can be obtained from the values for R_u and R_l by use of the following formula:

$$\text{Difficulty level (percentage passing)} = \frac{100\ (R_u + R_l)}{2f} = \frac{50(R_u + R_l)}{f} \qquad (3)$$

Since this is based on only the extreme fifty-four per cent of the cases, it does not give exactly the value for the index of item difficulty which would be obtained from all of the cases. In practice however the discrepancy usually is not appreciable.

Formula (1) for the U-L Index provides no specific correction for omissions, although it is necessary sometimes to item analyze speeded tests in which the number of omissions among later items is appreciable.

Omissions on speeded tests are usually of two types: (a) those by persons who read the item but fail to mark any choice and

(b) those by persons who have not read as far in the test as the end of that item. Davis* in a communication to the writer suggests a method of determining the 'non-reads' directly. The test papers are scored and arranged in rank order according to the criterion score. The group which includes the top twenty-seven per cent of papers is segregated as is the group which comprises the bottom twenty-seven per cent of test papers. Next the papers of the upper group are gone through to determine which item on each paper is the last one marked. A tally is then made opposite the number of the next item for that one is presumably the first item not read by the subject. This tally is the basis for a cumulative frequency table of the number of cases within the upper group who have not read each particular item. A similar count is made of the 'non-reads' in the lower group or the bottom twenty-seven per cent of the test papers.

By graphic item count on the International Test Scoring machine, or by other means, the number of right responses and the number of wrong responses in the upper group to each item is found. The omits may be obtained, if desired, by subtracting from the number of papers within the upper group the rights plus wrongs plus 'non-reads.' Similarly the omits among the lower group may be obtained, if desired.

Where the 'non-reads' represent an appreciable proportion of either criterion group the following formula for the U-L Index is suggested:

$$\frac{R_u}{(f - \text{`non-reads'}_u)} - \frac{R_l}{(f - \text{`non-reads'}_l)} \tag{4}$$

where R_u = number of right responses in the upper group
R_l = number of right responses in the lower group
$f = .27N$
'non-reads'$_u$ = 'non-reads' in the upper group
'non-reads'$_l$ = 'non-reads' in the lower group

The proportion $\frac{R_u}{(f - \text{`non-reads'}_u)}$ is the proportion of right responses in the upper group to the number in that group who read the item. The proportion $\frac{R_l}{(f - \text{`non-reads'}_l)}$ may be similarly interpreted for the lower group.

---

* Dr. Frederick B. Davis, formerly with the Coöperative Test Service and the Aviation Psychology Program of the Army Air Forces, now at Hunter College.

The U-L Index is proposed as a useful means of selecting items either to increase the internal consistency (odd-even reliability) of a test or to increase its correlation with a given criterion. Some advantages of the U-L Index are: (1) the demonstrated effectiveness of the upper vs. lower twenty-seven per cent technique on which it is based, (2) its ease of computation, and (3) the readiness with which its level of significance may be approximated.